PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2008) **363**, 923–945
doi:10.1098/rstb.2007.2151
Published online 7 September 2007

Neural representation of spectral and temporal information in speech

Eric D. Young*

Department of Biomedical Engineering, Centre for Hearing and Balance, Johns Hopkins University, 720 Rutland Avenue, Baltimore, MD 21205, USA

Speech is the most interesting and one of the most complex sounds dealt with by the auditory system. The neural representation of speech needs to capture those features of the signal on which the brain depends in language communication. Here we describe the representation of speech in the auditory nerve and in a few sites in the central nervous system from the perspective of the neural coding of important aspects of the signal. The representation is tonotopic, meaning that the speech signal is decomposed by frequency and different frequency components are represented in different populations of neurons. Essential to the representation are the properties of frequency tuning and nonlinear suppression. Tuning creates the decomposition of the signal by frequency, and nonlinear suppression is essential for maintaining the representation across sound levels. The representation changes in central auditory neurons by becoming more robust against changes in stimulus intensity and more transient. However, it is probable that the form of the representation at the auditory cortex is fundamentally different from that at lower levels, in that stimulus features other than the distribution of energy across frequency are analysed.

Keywords: auditory nerve; tonotopic; speech; discrimination; inferior colliculus; auditory cortex

1. INTRODUCTION

The general features of the neural representation of speech have been studied for over 25 years (e.g. Sachs & Young 1979; Young & Sachs 1979; Delgutte 1980; Reale & Geisler 1980; Sinex & Geisler 1983; Palmer *et al*. 1986). This is a challenging problem, in part owing to the complexity of the speech signal. Information in speech is encoded in a rapid sequence of different sound segments. The individual segments can be characterized by their frequency spectra, meaning the distribution across frequency of the energy making them up. The spectra change with the speech segment, so the resulting signal is a complex spectrotemporal pattern. (For the properties of speech, see the paper by Diehl (2008).)

This paper summarizes some aspects of the progress in understanding the representation of the speech signal in the brain, as reflected in the discharge patterns of single neurons and populations of neurons. The focus is on the peripheral parts of the auditory system, mainly the auditory nerve (AN), but results are shown also for three nuclei in the central auditory system, the cochlear nucleus, inferior colliculus and primary auditory cortex. Studies of responses to speech in auditory and other parts of the cortex using imaging and evoked electrical and magnetic signals are considered in other papers in this issue.

The problem of the neural representation of the spectrotemporal pattern of speech has generally been simplified either to the representation of the frequency spectrum of stationary speech segments or to the representation of the temporal sequence of energy within a narrow band of frequencies (those to which the neuron under study is sensitive). Here we consider the spectral representation first, then the representation of temporal stimulus patterns.

The frequency content of speech sounds is largely determined by the placement of the formant frequencies, which are the resonant frequencies of the vocal tract (Fant 1970). The speech signal has peaks of energy at the formant frequencies. For vowels, the formants dominate both the neural responses, described below, and the acoustic properties of the sound (Stevens & House 1961). Owing to this, much of this chapter is devoted to the representation of the formants. For consonants, the formants are still important; however, consonants often vary significantly with time, which means either that the formant frequencies vary with time (formant transitions) or that the sound contains periods of silence bordered by transients in sound energy, or both. Studies of the neural representation of consonants have often focused on stop consonants, in which both formant transitions and transients are prominent. Examples of those studies will be described.

The study of auditory neural representations begins from the assumption that the representation is tonotopic, meaning that the speech sound is decomposed into its component frequencies by the basilar membrane. Essentially, energy at different frequencies causes displacement of the basilar membrane at different frequency-specific locations (Robles & Ruggero 2001). Thus, the neural elements of the cochlea, the hair cells and AN fibres, sense energy at different frequencies, depending on their location along the basilar membrane. As a result, different features of a

*eyoung@jhu.edu

One contribution of 13 to a Theme Issue 'The perception of speech: from sound to meaning'.

speech sound, such as the formants, are represented in separate populations of AN fibres, according to their frequencies. Presumably, this separation is important in speech perception because it minimizes the interference between speech components at different frequencies. For example, the second formant (F2) can be masked by the generally more intense first formant (F1); this effect is larger in persons with impaired hearing (reviewed by Moore 1995), where the separation of the formant representations breaks down (Miller *et al.* 1997). Thus, the fact that an auditory neuron responds only to a narrow range of frequencies, called tuning, is the most important factor in the neural representation of speech. However, other properties of cochlear transduction are also important, especially suppression of the responses at one stimulus frequency by energy at another. These points will be illustrated by examples.

For those unfamiliar with neurophysiological studies of the auditory system, a general description of the experimental basis for the results discussed in this paper is given in appendix A.

2. TUNING IN THE AUDITORY NERVE: THE TONOTOPIC REPRESENTATION

The basic representation of stimuli at all levels of the auditory system is tonotopic (Schreiner *et al.* 2000), meaning that different frequencies in the stimulus are analysed separately (although not independently). The initial frequency analysis occurs in the cochlea. The AN conveys to the brain a representation of sound that resembles the outputs of a bank of bandpass filters tuned to different frequencies. For AN fibres, the tuning can be described in two ways as shown in figure 1. Figure 1*a* shows tuning curves for 11 AN fibres from one cat. Each curve shows the threshold sound level of a fibre plotted against the stimulus frequency; threshold here means the sound level producing a criterion increase in discharge rate (usually 20 spikes s^{-1}). Each curve is V-shaped with a minimum threshold at the best frequency (BF) of the fibre.

It is clear that each fibre has a restricted range of frequencies to which it responds and that there are fibres tuned to frequencies (i.e. with BFs) across the audible range of the animal. Assuming that the tuning curves are the (inverted) gain functions of bandpass filters and that the filters are an accurate model of the neurons' input/output characteristics, one expects that each fibre should convey information to the brain about stimulus frequencies near the fibre's BF, thus decomposing the stimulus by frequency. The assumption that such filters are an accurate model for AN fibres is only partly true, in that there are nonlinear effects that complicate cochlear tuning. Most important are the nonlinear interactions, which allow energy at one frequency to suppress responses to a different frequency (Sachs 1969; Javel *et al.* 1978; Javel 1981; Delgutte 1990).

A more accurate measure of suprathreshold frequency selectivity is provided by the gain functions in figure 1*b*, for a chinchilla AN fibre tuned to 1.35 kHz (Recio-Spinoso *et al.* 2005). Unlike the tuning curves, these are true gain functions, with units of response divided by stimulus amplitude. The gain curves were

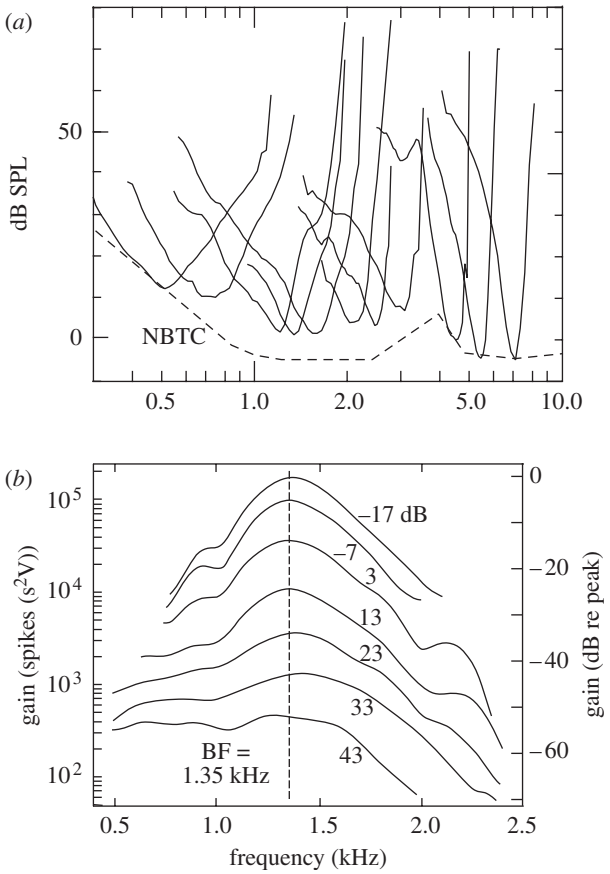

Figure 1. (*a*) Tuning curves of AN fibres from a cat for BFs up to 7 kHz. The dashed line marked NBTC shows the thresholds of the most sensitive fibres in a group of cats. (*b*) Gain versus frequency functions for an AN fibre measured at different sound levels using the method of reverse correlation. The stimulus was a broadband noise and the gain functions express the discharge rate of the neuron divided by the power spectral density of the stimulus. Gain is given in absolute units on the left ordinate and as dB relative to the peak gain on the right ordinate. The units of gain on the left ordinate derive from the way the gain functions are computed (see Johnson (1980*a*) for a full explanation). The sound levels of the noise are given as spectrum levels next to the curves, as dB re 20 μPa $Hz^{-(1/2)}$, the level in a 1 Hz wide band. The speech sound levels given in subsequent figures are not comparable with these spectrum levels, because the speech levels are the overall intensity of the stimulus, summed across frequency. The vertical dashed line is the approximate BF. (*a*) Adapted with permission from Miller *et al.* (1999*a*) and (*b*) adapted with permission from Recio-Spinoso *et al.* (2005).

computed from the responses to broadband Gaussian noise as the first Wiener kernel (Johnson 1980*a*), a method similar to reverse correlation (de Boer & de Jongh 1978). Briefly, the data are the responses of the neuron to a broadband noise at the ear. The average waveform of the noise preceding each action potential is computed, resulting in the reverse correlation or revcor function (Møller 1977; de Boer & de Jongh 1978; Carney & Yin 1988; Lewis & Henry 1994; Recio-Spinoso *et al.* 2005). The Fourier transform of the revcor is the gain function of the linear filter that best approximates the input/output characteristics of the neuron, for the stimulus conditions used to obtain the data. Figure 1*b* shows the gain functions at seven sound levels for this fibre; note that the shape of the

gain functions changes with the sound level. Thus, at low noise levels (−17 dB), the gain function is bandpass and is very close to the inverted tuning curve of the neuron; at high noise levels (43 dB), the gain function is broad and low-pass. This behaviour is similar to that of basilar membrane gain functions (Ruggero *et al.* 1997) and is the basis for models of AN responses that successfully model responses to speech (Carney 1993; Bruce *et al.* 2003; Holmes *et al.* 2004).

The gain functions in figure 1*b* show that AN fibres are less frequency selective at high sound levels. This suggests that the neural representation of broadband stimuli such as speech should change with sound level. For example, for a vowel, the energy at F1 is typically 10–20 dB more intense than the energy at F2. Thus for neurons with BFs near F2, sharp bandpass filtering is necessary if the neuron is to respond to F2 in preference to F1. In figure 1*b*, if a vowel were presented with F2 at the neuron's BF and F1 an octave lower, the neuron would respond to F2 at low sound levels, where the filtering is sufficiently sharp to attenuate F1 below F2, but would respond to F1 at higher sound levels where the filtering is low-pass and does not attenuate F1 relative to F2. This behaviour is observed experimentally, in that AN fibres with BFs near F2 gradually switch their responses from F2 to F1 at sound levels between 70 and 90 dB SPL (Wong *et al.* 1998).

3. ANALYSIS OF THE NEURAL REPRESENTATION

Figure 2 shows an example of the responses of an AN fibre to speech and illustrates the way in which neural representations are constructed from the data. The waveform of the sentence 'five women played basketball' is shown at low time resolution in figure 2*a*. The peri-stimulus time (PST) histogram in figure 2*b* shows the rate of discharge of the fibre in response to the speech, on the same time axis. This histogram was constructed by counting spikes in bins synchronized with the stimulus. From the PST histogram, it is possible to estimate the strength of response of the neuron to the stimulus (or to any temporal segment of the stimulus) as the average rate of spiking over the appropriate time interval (number of spikes/length of time interval). Clearly, the different syllables activate this neuron to different degrees, presumably because they differ in the amount of energy in the neuron's tuning curve.

A rate representation could be constructed from data like these, obtained from a population of neurons with different BFs, by computing the average discharge rate of each neuron during the stimulus segment of interest and plotting those rates versus the BFs of the neurons. For example, the average rate over 0.1–0.23 s could be computed to obtain rate responses to the vowel nucleus in 'five'. Rate representations have generally been computed with low time resolution, over time intervals of 100 ms or more, usually from responses to a stationary stimulus like a steady vowel. However, rate representations can be computed at any time resolution from responses to any stimulus.

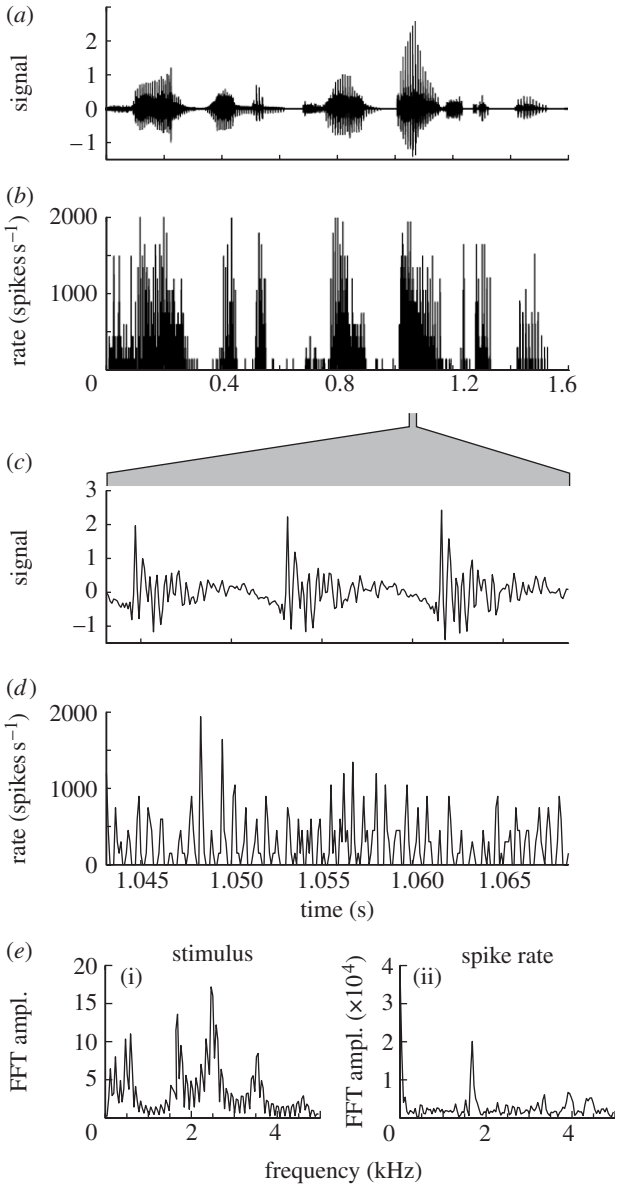

Figure 2. (*a*) The waveform of the sentence 'Five women played basketball' is shown at low time resolution. (*b*) A PST histogram of the response of an AN fibre (BF = 1.7 kHz) to the stimulus in (*a*). The PST histogram was constructed by counting the number of spikes that occurred during successive 0.1 ms bins synchronized with the stimulus, over 67 repeats of the stimulus. Although the PST histogram was computed at high resolution, it is displayed at low resolution, so only the overall changes in discharge rate can be seen. (*c*,*d*). These are the same plots as in (*a*,*b*), except at higher time resolution. The portions of the low-resolution plots between 1.042 and 1.068 s are shown, as indicated by the schematic between (*b*,*c*). (*e*(i)) Magnitude of the discrete Fourier transform of the stimulus segment in (*c*); (*e*(ii)) same for the neural response in (*d*). The major peaks in (*e*(i)) correspond to F1 (near 0.5 kHz), F2 (near 1.7 kHz), F3 (near 2.5 kHz) and F4 (near 3.5 kHz); the response plot (*e*(ii)) shows only a single peak near 1.7 kHz. Note that both plots have a linear ordinate.

Examples will be shown in figure 5 for 200 ms and figure 7 for 1 ms resolution.

In the rate representation discussed above, there is no way to know which components of the stimulus (e.g. F1 versus F2) produced the response. It is possible to gain further information by looking at higher time

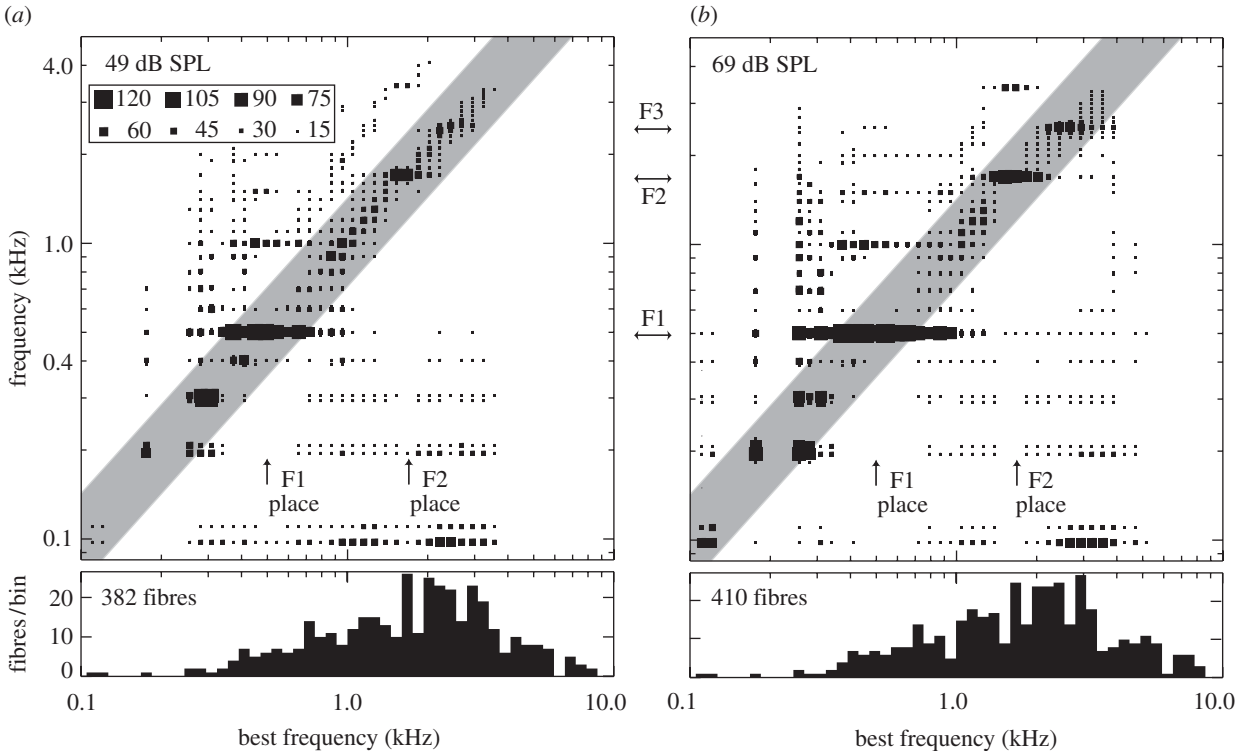

Figure 3. (*a*,*b*) Three-dimensional plots of the analysis of phase-locking in a population of AN fibres responding to the vowel /ɛ/. Responses to two sound levels are shown. The abscissae show fibre BFs and the ordinates show the frequencies to which the fibres are phase-locked. The spectrum of the vowel is shown in figure 5*a* (F2 = 1.7 kHz) and the formant frequencies are indicated by the horizontal double-headed arrows between the plots. The formant frequencies are also shown along the BF axis by the vertical arrows. The response strength is indicated by the size of the box, using the scale shown in the inset of (*a*). For example, rates greater than 120 spikes s^{-1} are shown by the largest box, rates between 105 and 120 spikes s^{-1} by the second largest box, etc. The responses are magnitudes of discrete Fourier transforms of PST histograms, as in figure 2*e*, averaged across all fibres with BFs within a frequency bin (0.133 octave). The grey stripe shows where the frequency of the response is within 0.5 octave of BF. Thus, responses within the stripe are tonotopic. The number of fibres in each bin is shown by the histograms below the plots. For this analysis, fibres were combined across ten normal animals with similar threshold audiograms. Adapted with permission from Schilling *et al.* (1998).

resolution, as in figure 2*c*,*d*. The high-resolution rate plot in figure 2*d* shows that the neuron's activity is not random, but strongly periodic at a frequency of approximately 1700 Hz. Although it is not obvious in this figure, the 1700 Hz oscillation is time-locked to a similar oscillation in the stimulus; this is an example of the well-known phase-locking property of auditory neurons (Rose *et al.* 1967; Young & Sachs 1979; Johnson 1980*b*; Palmer & Russell 1986), in which their spikes occur at a particular phase of the cycle of a periodic stimulus.

By analysing the frequency components that are present in the response and locked to the stimulus, it is possible to infer which components of the stimulus are effective in exciting the neuron. This analysis is shown in figure 2*e* which shows the magnitude of the Fourier transform of the stimulus in figure 2*e*(i) and the magnitude of the transform of the spike rate in figure 2*e*(ii); these are the transforms of the signals in figure 2*c*,*d*, respectively. Whereas the stimulus contains many significant frequency components, the response is mainly to frequencies near 1.7 kHz, the F2 frequency during this part of the stimulus. Of course, this reflects the tuning of the neuron. On the basis of this analysis, one concludes that this neuron is responding to the F2 energy in the stimulus and therefore 'represents' F2.

4. TONOTOPIC REPRESENTATION OF VOWELS

AN responses to vowels are usually dominated by the formant frequencies (Young & Sachs 1979; Delgutte 1984; Delgutte & Kiang 1984*c*; Sinex & Geisler 1984; Palmer *et al.* 1986). Figure 3 shows an example for responses to the vowel /ɛ/ as in 'met' at two sound levels (Schilling *et al.* 1998). The method of figure 2 has been used to analyse the components of the response based on phase locking. The data consist of the responses to the vowel in a population of several hundred AN fibres. The fibres have been sorted into bins according to BF, shown along the abscissae of the plots. The Fourier transforms of the responses were computed for each fibre and the magnitudes of the transforms were averaged for all the fibres falling into each bin. The box plots show the averages as a function of frequency, plotted along the ordinate. Thus, the plot shows the tonotopic array of neurons along the abscissa and the strength of the response to various frequency components of the vowel along the ordinate.

Note the concentration of larger boxes at the frequencies of the formants along the ordinate. The largest responses are to F1 (0.5 kHz), with smaller responses to F2 (1.7 kHz) and F3 (2.5 kHz). These responses are centred (along the abscissa) on fibres with BFs near the frequency of the appropriate formant. There are also significant responses to the

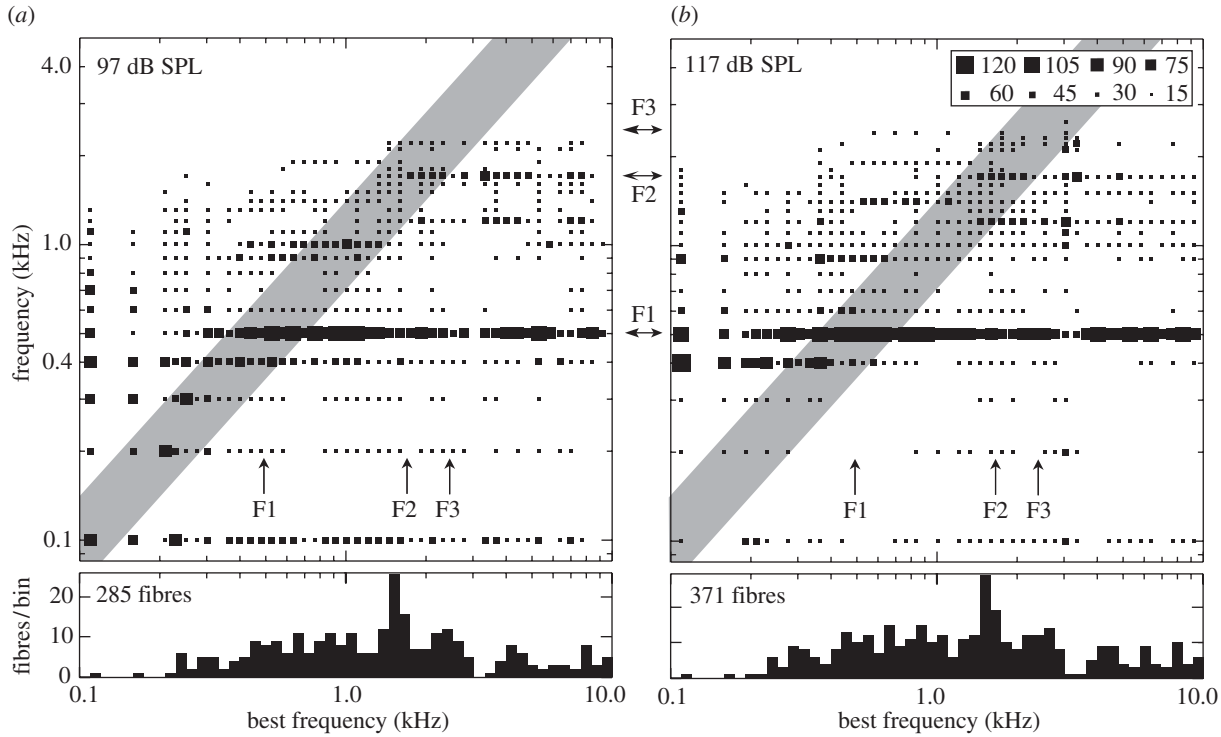

Figure 4. (*a*,*b*) Box plots of population responses to the vowel /ɛ/ at two sound levels, as in figure 3. The data are from four animals exposed to acoustic trauma. The trauma produced a threshold elevation of approximately 20 dB for BFs up to 1 kHz and 40–60 dB at higher BFs (see fig. 2 of Miller *et al.* (1997)). The sound levels shown are about the same number of dB above threshold as for figure 3 for BFs near F2 (1.7 kHz).

fundamental frequency of the vowel (100 Hz) among fibres with higher BFs. These neurons have several frequency components of the vowel within their tuning curves, and their BFs are high enough that phase locking to those near-BF frequency components is weak; as a result, they respond significantly to the envelope of the sum of those components, at the fundamental frequency of the vowel. Responses to other frequency components of the vowel are smaller than those to the formants, except for a response to the second harmonic of F1 (1 kHz). This response has previously been discussed and is caused by a rectification artefact in the analysis (Young & Sachs 1979); it should be considered as part of the F1 response.

As the sound level increases from 49 to 69 dB SPL, the responses to the formants increase in magnitude and spread along the BF axis to occupy a larger fraction of the population. This spread is qualitatively consistent with the broadening of fibres' gain functions in figure 1*b*.

The earlier statement that AN responses to speech are tonotopic is illustrated by the fact that the responses in figure 3 are mostly within the grey stripes. These stripes show where response frequency is near BF (within 0.5 octave). Most important, the population of neurons responding to F1 is different from the population responding to F2 and F3. Two mechanisms are important in maintaining the tonotopic representation: (i) basilar membrane tuning is essential to separate the components initially (Geisler 1989) and (ii) suppression sharpens the representation by restricting the spread of components along the BF axis.

The effect of suppression can be seen in the 69 dB SPL data in figure 3 in that the F1 response amplitude decreases sharply at BFs close to F2. At high sound levels, the responses to F1 and F2 are reciprocal, in that at BFs where the F2 response is large, the F1 response is small. That the reciprocal behaviour is suppression follows from the observation that a tone at the level and frequency of F1 in the 69 dB SPL data in figure 3, presented by itself, would produce strong responses phase-locked to the F1 frequency among fibres with BFs near F2 (Wong *et al.* 1998). The fact that those responses are not seen in figure 3 implies that they are suppressed by F2 (for a similar analysis using two tones, see Kim *et al.* (1979)). Further support for the importance of suppression comes from the fact that a linear model of AN fibres that does not include suppression does not display the kind of reciprocal responses observed in real fibres (Sinex & Geisler 1984), whereas a nonlinear model with suppression does (Deng & Geisler 1987*a*).

The importance of the cochlear mechanisms that maintain the normal tonotopic representation can be seen by looking at responses in the AN when the cochlea has been damaged by acoustic trauma (Miller *et al.* 1997). For example, in figure 4, acoustic trauma was produced by exposing the animals to an intense narrow band of noise centred at 2 kHz for several hours (Miller *et al.* 1999*a*). The exposure produced a high-frequency hearing loss, in which the thresholds and tuning curves were within 20 dB of normal for BFs up to approximately 1 kHz; at higher BFs, there was significant threshold elevation, between 40 and 60 dB, and tuning curves were abnormal by being broader than usual. In addition, the strength of suppression was decreased, as measured using two-tone interactions (Schmiedt *et al.* 1980; Salvi *et al.* 1982; Miller *et al.* 1997). When the cochleae of animals given similar

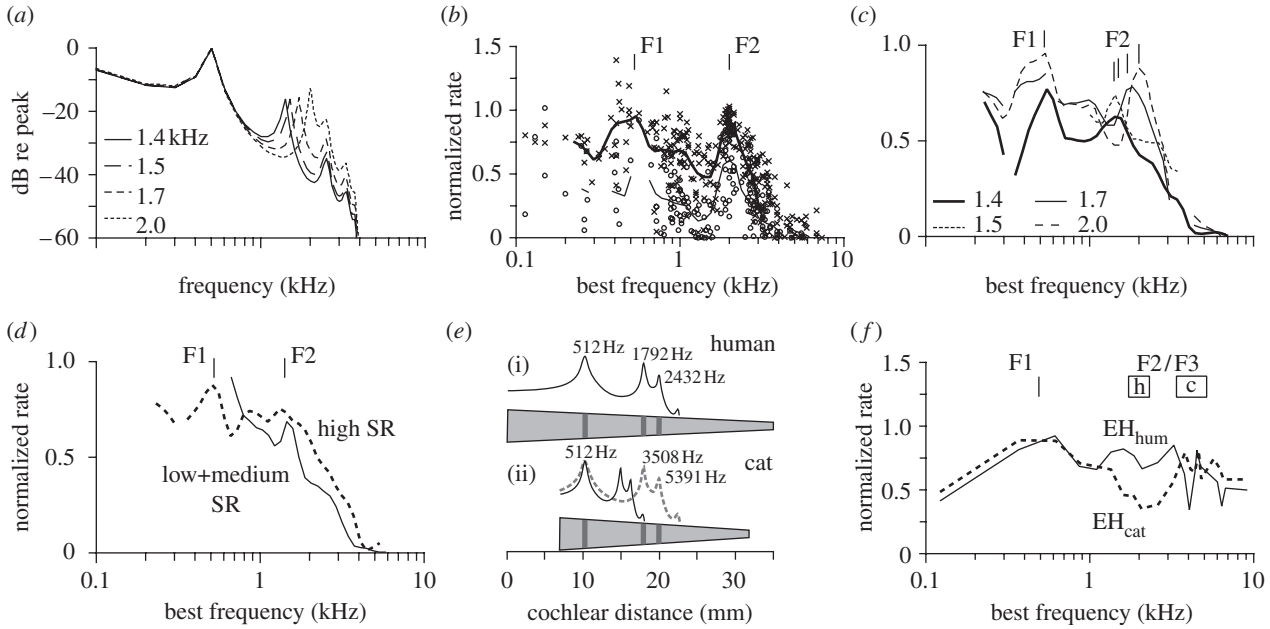

Figure 5. Rate profiles of responses to vowels similar to /ɛ/. (*a*) Spectra of four variants of /ɛ/ with the same F1 and F3, but different F2 frequencies. For all variants, F1 = 0.5 kHz and F3 = 2.5 kHz. (*b*) Empirical rate–place representation of the F2 = 2 kHz variant at 50 dB SPL (open circles) and 70 dB SPL (crosses), computed from 200 ms stimulus presentations repeated 50 times. Each point shows the discharge rate of a fibre plotted against its BF. Rate is normalized as (rate − SR)/(maxrate − SR), where maxrate is the rate in response to a BF tone 50 dB above threshold and SR is spontaneous rate. The lines are moving-window averages computed from the data points with a log-triangular window of width 0.15 decade. The heavy and light lines are averages for 70 and 50 dB SPL, respectively. Lines are not plotted for windows with fewer than 1.5 fibres. The formant frequencies are marked at the top of the plot. (*c*) Moving-window averages of normalized rate profiles for the four variants of /ɛ/ at 70 dB SPL. The F2 frequencies are given. The data are sparse at BFs near F1. (*d*) Normalized rate profiles for high SR fibres (SR ≥ 20 spikes s^{-1}) and low + medium SR (LMSR) fibres (SR < 20 spikes s^{-1}) in response to the F2 = 1.7 kHz variant at 70 dB SPL. (*e*) Schematic comparison of the position of the spectra of /ɛ/ on the (i) human and (ii) cat basilar membranes. The vowel spectra are plotted with an abscissa that matches the layout of frequencies along the basilar membrane. The formant frequencies are indicated next to the spectra. For the cat, the spectrum is shown twice, once (black) with the usual formant frequencies and once (grey, dashed) with the spectrum spread out to make the physical distance between the formants the same on the cat and human basilar membrane. The schematics show the positions of the formants on the two membranes, for the normal /ɛ/ on the human membrane and the spread /ɛ/ on the cat membrane. (*f*) Normalized rate profiles for /ɛ/ with its standard formants (EH_{hum}) and with the formants spread as in (*e*) (EH_{cat}). The positions of the formants are marked at the top of the plot; F2 and F3 frequencies are shown as the left and right sides of the small boxes labelled 'h' for human and 'c' for cat. (*a–d*) Adapted with permission from Conley & Keilson (1995), (*e*) adapted with permission from Kiefte & Kluender (2001) and (*f*) adapted with permission from Recio *et al.* (2002).

exposures were examined anatomically, there was a partial loss of outer and inner hair cells and damage to the transduction apparatus in the remaining hair cells (Liberman & Beil 1979; Liberman & Dodds 1984).

The broadening of tuning and decrease in suppression both reduce the separation of responses to F1 and F2, with the result that the representation of the vowel is degraded (Palmer 1990; Palmer & Moorjani 1993; Miller *et al.* 1997). Two changes are evident in the example in figure 4, in comparison with figure 3. First, responses to F1 dominate the responses of fibres at all BFs tested. Responses to F2 and F3 are hardly seen. This is expected from the degraded tuning, because the broadened tuning curves now do not adequately discriminate between F1 and F2 in neurons with BFs near F2; this is similar to the effect of degraded tuning at high sound levels in figure 1*b*. Second, there is a wide distribution along the ordinate of phase-locking to frequencies other than the formants, regardless of BF. This broadband phase-locking is consistent with decreased suppression, in the sense that the fibres' responses are not captured by the strong responses to the formants, as they would be in a normal ear.

5. RATE REPRESENTATION OF SPECTRAL SHAPE

The results in figure 3 show that fibres normally respond most strongly for BFs near the formants. This result leads to the simplest neural representation of spectral shape, the *rate–place profile* (Sachs & Young 1979; Delgutte & Kiang 1984*b*), a plot of discharge rate versus BF in the population of AN fibres. Such a profile should have a shape similar to the spectrum of the stimulus. However, there are substantial responses at BFs between the formants in figure 3, and the spread of response along the BF axis as sound level increases suggests that the response should become more uniform across BF at high sound levels. Thus, it is not clear how well a rate–place profile will convey information about the stimulus spectrum. The data in figures 5 and 6 will be used to explore this question.

Figure 5*a* shows the spectra of four variants of the vowel /ɛ/ with different F2 frequencies. The question posed in the previous paragraph will be answered here by considering how well the responses to these four stimuli can be differentiated. Rate profiles for a population of AN fibres responding to the F2 = 2 kHz

variant are shown in figure 5*b* (Conley & Keilson 1995); rates are shown for two sound levels (50 and 70 dB SPL), indicated by the different symbols and line weights. The data points show the responses of individual fibres and the lines show moving-window averages of the data points. Clear rate peaks are seen among fibres with BFs equal to F2 at both sound levels and at BFs near F1 at the higher level; presumably a rate peak at F1 would have been observed at the lower level with sufficient data. No separate peak at F3 is evident. The population response can be said to represent the spectrum of the vowel in the sense that the first two formant frequencies can be estimated from the BF locations of the rate peaks.

An informal measure of the quality of the representation is how well the rate peak at F2 stands out, as measured by its height, the difference in rate between BFs near F2 and the rate minimum for BFs between the formants. The peak height is slightly greater at 50 than at 70 dB SPL in figure 5*b*. This trend, towards poorer rate representation at higher sound levels, has been reported for several vowels (Sachs & Young 1979).

The F2 peak height depends on the F2 frequency, increasing as F2 moves away from F1. Figure 5*c* shows moving-window averages for the 70 dB SPL stimuli for all four variants of the vowel. For the lowest F2 frequency (1.4 kHz, heavy line), there is only a small rate peak; the size of the rate peak grows, both in absolute and relative terms, as the F2 frequency increases. The increase in the rate peak as the separation between the frequencies of F1 and F2 increases is probably attributable to a reduction in suppression by F1 of the response to F2 as the frequencies move farther apart. The argument for this point is that the rate of neurons with BFs near F2 changes with sound level of the vowel in the same way as rate in response to a tone at BF = F2 in a two-tone suppression paradigm, with the second tone at F1 (Sachs & Young 1979, fig. 13).

Thus, suppression has two countervailing roles: suppression of responses to F1 at the F2 place is important in maintaining the tonotopic representation at high sound levels, but suppression of responses to F2 by F1 reduces the salience of rate peaks in response to F2, as F2 approaches F1. These data suggest that the rate representation of F2 will vary with the vowel and will be stronger in vowels with more separation between the formants.

The peak height also varies in different groups of AN fibres. Fibres vary in their thresholds and dynamic ranges, properties that are correlated with spontaneous discharge rate (SR; Liberman 1978) and the mode of innervation of hair cells (Liberman 1980). Dynamic range means the range of sound levels over which the fibre changes its rate when the input changes in level. Low and medium SR fibres generally have higher thresholds and wider dynamic ranges and provide a better rate representation at high sound levels (Sachs & Abbas 1974; Sachs & Young 1979; Winter *et al.* 1990; Yates *et al.* 1990). Figure 5*d* compares the rate representation of the F2 = 1.7 kHz variant of the vowel between fibres with SR < 20 spikes s^{-1} ('low + medium SR') and fibres with SR ≥ 20 spikes s^{-1} ('high SR'). A small peak is observed for the low and medium SR fibres, but not for the high SR fibres, suggesting that the rate peaks in figure 5*c* result mainly from activity in the low and medium SR populations.

6. THE QUESTION OF WHETHER THE CAT IS A GOOD MODEL FOR THE HUMAN EAR

An important question about the data shown here is whether there are systematic differences between the human cochlea and the cochlea of whatever animal model is being used and whether those differences are important to studies of speech coding. In simplest terms, one wonders to what extent data like those in figures 3–5 need to be corrected in some way before they apply to the neural representation of speech in the human ear. This question has a number of facets, such as the use of anaesthesia in animal experiments, the effects of damage done to neural circuits by the surgery necessary for neurophysiological recording and the possible differences between the auditory systems of a species that uses language and others that do not. In this section we will focus on another question, whether there are differences in frequency representation between human and animal auditory systems.

Although the general properties of cochlear physiology are similar across mammals, there are differences in the frequency range of hearing and the length of the cochlea (Fay 1988; Greenwood 1990). For example, human hearing extends from approximately 20 to 15 kHz in a 35 mm cochlea and cat hearing extends from approximately 90 to 60 kHz in a 25 mm cochlea. These differences raise the question of possible differences in the frequency resolution of the cochlea across species. The cochlear frequency maps of laboratory animals (the map of BF onto distance along the basilar membrane) have the frequencies closer together, compared with the human ear. The extent of crowding for the vowel /ɛ/ is shown in figure 5*e*, which shows a schematic of the spectrum of /ɛ/ (black solid lines) mapped onto the basilar membrane of the human (above) and the cat cochlea (below; Kiefte & Kluender 2001). The spacing between the F1 and F2 is smaller by a factor of approximately 0.6 on the cat, versus the human, basilar membrane.

In order to get an idea of the potential effects of the difference in cochlear maps, Recio *et al.* (2002) assumed that interactions between different frequencies occur over a constant physical distance along the basilar membrane in different species. There is no evidence for this assumption, but it is consistent with the fact that auditory filter bandwidths correspond to roughly a constant distance along the basilar membrane within a species (Greenwood 1961, 1990; see the paper by Moore 2008). Recio and colleagues synthesized a version of /ɛ/ with modified formant frequencies, so that the formants were the same physical distance apart in the cat cochlea as they are in the human cochlea. The spectrum of the modified vowel is shown by the grey dashed line in figure 5*e*. Figure 5*f* compares rate profiles from cat AN fibres for two vowels: EH_{hum} is the standard /ɛ/ and EH_{cat} is the modified version. The rate profile obtained with EH_{hum} (solid line) is similar to the profiles in figure 5*c*,*d*, with a small peak at F2 and no sign of a separate peak for F3.

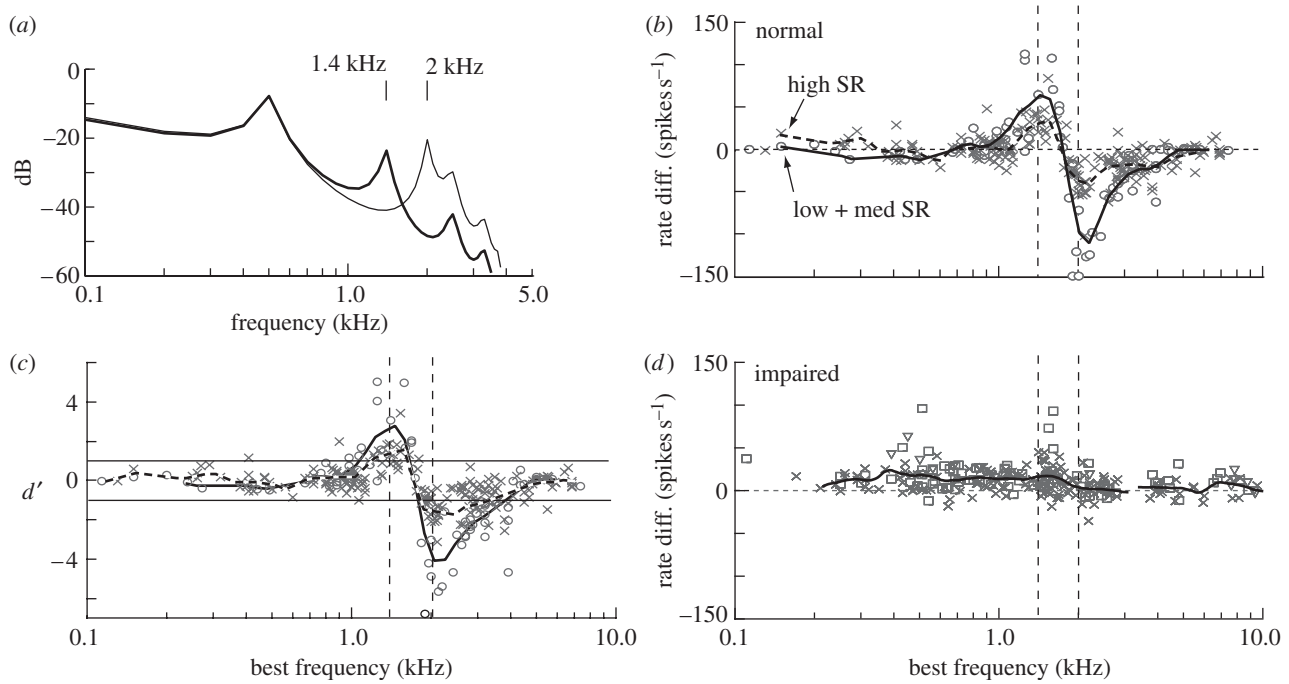

Figure 6. (*a*) The spectra of two variants of the vowel /ɛ/ with F2 frequencies of 1.4 and 2 kHz (the same as in figure 5*a*). (*b*) Rate difference between responses to the two variants presented at 70 dB SPL (rate to the F2 = 1.4 kHz variant − rate to the F2 = 2 kHz variant). Data are from the same population as in figure 5*b–d*, and rate differences are plotted against BF. SRs are identified by symbol: open circles and solid lines are for SR < 20 spikes s^{-1}, crosses and dashed lines are for SR ≥ 20 spikes s^{-1}. Lines are moving-window averages as in figure 5. Vertical dashed lines show the two F2 frequencies. (*c*) The data of (*b*) replotted as d' versus BF. The horizontal lines show $d' = 1$. (*d*) Rate differences for the same stimuli from a population of AN fibres studied in cats with a high-frequency threshold shift due to acoustic trauma, the same population as in figure 4. The plot is the same as (*b*) except that triangles are for SR < 1 spikes s^{-1} and squares are for SR between 1 and 20 spikes s^{-1}. The moving-window average is for all SRs. The stimulus level was 97 dB SPL. (*a*,*d*) Adapted with permission from Miller *et al.* (1999*b*), (*b*,*c*) adapted with permission from Conley & Keilson (1995).

The rate profile for EH_{cat} (dashed line) shows clearly defined rate peaks for F2 and F3, and a substantial improvement in the F2 peak height. The increase in peak height is mainly due to a decrease in normalized rate at the rate minimum between F1 and F2 (near 2 kHz). This change is probably due to the narrower relative filter widths (relative to the stimulus spectrum) because it is not consistent with weaker suppression of more widely spaced frequency components.

Evaluation of the differences among species has mainly focused on the width of tuning (reviewed by Ruggero & Temchin 2005). Unfortunately, only psychophysical or indirect physiological measures, such as the compound-action-potential tuning curve (Eggermont 1977), can be obtained in human subjects, which raises a problem of interpretation. The width of an AN tuning curve is a straightforward measure, but the results for psychophysical or indirect physiological filters depend on the methods used to obtain the data (reviewed by Moore 2003). Presumably, these technical problems explain why some studies have obtained psychophysical or indirect physiological filters that match the AN tuning curves (e.g. Evans 1992), whereas others have not (e.g. Pickles 1979; Harrison *et al.* 1981). Thus, while most studies have found behavioural filters to be wider in animals than in human subjects, it is not clear how that result should be interpreted in terms of the relative widths of AN tuning curves (Ruggero & Temchin 2005).

Recently, Shera *et al.* (2002) published a direct comparison of human, cat and guinea-pig basilar membrane tuning, suggesting that tuning is at least twice as sharp in the human as in the cat or guinea-pig cochlea. They inferred tuning width from the group delay of otoacoustic emissions, a measurement that was done in both human and animal subjects. The results were supported by comparable psychophysical measures of tuning in the human cochlea that were done with forward masking, which they argued gives the most accurate measure of bandwidth (Oxenham & Shera 2003). However, Siegel *et al.* (2005) have argued that the assumptions underlying the analysis of otoacoustic emissions by Shera *et al.* are incorrect, based on basilar membrane measurements in the chinchilla cochlea. Moreover, Ruggero & Temchin (2005) have conducted a meta-analysis of behavioural and physiological data from a variety of species and concluded that forward masking underestimates the widths of physiological tuning and that human tuning curve widths are probably comparable with those in common laboratory animals. It is not clear which of the views summarized above is correct. Nevertheless, the current evidence does not unequivocally support a difference in tuning between the human cochlea and that of common laboratory animals. This argument applies to tuning; a similar analysis of suppression has not been done.

7. INFERENCES FROM DISCRIMINATION DATA

The data in figure 5 suggest that the variants of the vowel /ɛ/ whose spectra are shown in figure 5*a* would easily be discriminated on the basis of AN discharge

rates (and in fact they are easily discriminated in informal listening tests). However, the scatter in the discharge rates (figure 5*b*) and the small size of the rate differences as F2 approaches F1 (figure 5*c*) suggest that it would be profitable to quantify this impression. One approach is to measure the change in discharge rate as F2 changes and to compute from the rate changes the detectability of changes in F2 frequency for various stimuli (Conley & Keilson 1995; May *et al.* 1996; Miller *et al.* 1999*b*). Figure 6*b* shows rate differences between responses to variants of /ε/ with F2 = 1.4 and 2 kHz, plotted versus BF. The points are data from individual fibres and the lines are moving-window averages of the points.

The largest rate changes are in fibres with BFs equal to the two F2 frequencies (the vertical dashed lines). The rate changes are larger in fibres with low and medium SRs (open symbols), consistent with figure 5*d*. The fact that rate changes are largest in fibres with BFs equal to the formant frequencies is expected from the tonotopic representation of the vowel (figure 3) and provides direct evidence for the statement that these neurons represent F2.

The rate changes provide a reliable representation of F2 frequency to the extent that the rate changes themselves are reliably detectable. Because AN fibres give randomly varying rates, this is a statistical problem; detectability can be measured as the rate change divided by the standard deviation of the rates (Green & Swets 1966), called d'; Conley & Keilson (1995) computed d' as $(\mu_1 - \mu_2)/(\sigma_1^2 + \sigma_2^2)^{1/2}$, where μ_j and σ_j are the mean and s.d., respectively, of the rate in response to vowel j. The d' results in figure 6*c* show that rate changes in the best fibres (those with BFs equal to the F2 frequencies) easily achieve a d' value of 1, usually taken as the detectability at the discrimination threshold (or 'jnd' for just-noticeable difference).

The data in figure 6*b*,*c* are for the largest F2 difference in the dataset, between vowels with F2 frequencies of 1.4 and 2 kHz. Using data from the other vowels studied, Conley & Keilson (1995) showed that the threshold for discriminating the vowels, where $d' = 1$ at the peak of the moving-window average of the rate difference, occurred for a change in F2 frequency between 125 and 240 Hz, depending on SR. Vowel formant discrimination thresholds for F2 are usually 50–100 Hz in human observers depending on vowel and formant frequency (e.g. Liu & Kewley-Port 2004). The value computed by Conley & Keilson is based on information from a single average AN fibre. If information were combined across different fibres, the jnd would be considerably smaller (approx. 1 Hz). Thus, the psychophysical performance on vowel formant discrimination could easily be achieved from rate responses of AN fibres.

To emphasize the point made in figure 4, that the representation of F2 is lost following acoustic trauma, figure 6*d* shows rate differences for the population of impaired fibres from figure 4. There is minimal rate change related to F2 in this population. The reason, of course, is that the fibres are mostly responding to F1, which does not change between these two stimuli.

May *et al.* (1996) obtained a similar result for discrimination of the F2 frequency using a different method in which a model was fitted to the rate responses to vowels and d' was computed from the model. The model is based on data like those to be discussed later in figure 9. They also measured the behavioural jnd for F2 frequency in cats (Hienz *et al.* 1996; May *et al.* 1996), and showed that the jnd predicted for one optimally chosen AN fibre is very close to the behavioural jnd. 'Optimally chosen' means a fibre with a BF at the peak of the rate difference plot. At low sound levels and in quiet, high SR fibres provide the best information (such stimuli may be below threshold for low and medium SR fibres), whereas at high levels and in background noise, low and medium SR fibres provide the best information, because high SR fibres' rate responses saturate.

8. RESPONSES TO CONSONANT–VOWEL SYLLABLES

The stimuli considered so far are not typical of real-world speech in that they are isolated vowels, with spectra that are constant in time. The study of more realistic stimuli began almost as early as work on vowels, with studies that focused on stop–consonant–vowel stimuli (Miller & Sachs 1983; Sinex & Geisler 1983; Carney & Geisler 1986). Stops at the beginning of syllables are characterized by a release burst of broadband noisy sound followed after some delay by the onset of voicing; the voicing produces a vowel-like formant structure in which the formants move rapidly (over a few tens of ms) from frequencies determined by the consonant to frequencies appropriate for the following vowel.

Studies of the neural representation of stops indicated that the principles described above for vowels could be extended to consonants using both rate and phase-locked representations that varied in time. In these analyses, the spectrum of the stimulus was considered to be a sequence of slightly different spectra in successive time windows. The rate or phase-locked representation was computed separately in each time window, with results similar to those in figures 3 and 5. In particular, the responses were dominated by the formants, either through peaks of discharge rate at BFs that tracked the formants or through phase-locking that was dominated by the formant frequencies. Generally, the rate representation was found to be better for stop consonants than for vowels, probably because the consonants have lower sound levels. Responses to other consonants, including nasals and fricatives, have also been analysed, with similar results (Delgutte & Kiang 1984*b*; Deng & Geisler 1987*b*).

The approach of using discriminability as a measure of the quality of a neural representation can provide additional insight into the representation of consonant–vowel (CV) syllables. Figure 7*a* shows differences between the rates of AN fibres in response to the synthetic utterances /bab ba/ and /dad da/ (Bandyopadhyay & Young 2004). The data are from a population of fibres, arranged along the ordinate according to BF. The abscissa shows time during the stimulus, and the components of the stimulus are marked above the plot. The first three formant frequencies of the two stimuli are shown by the black

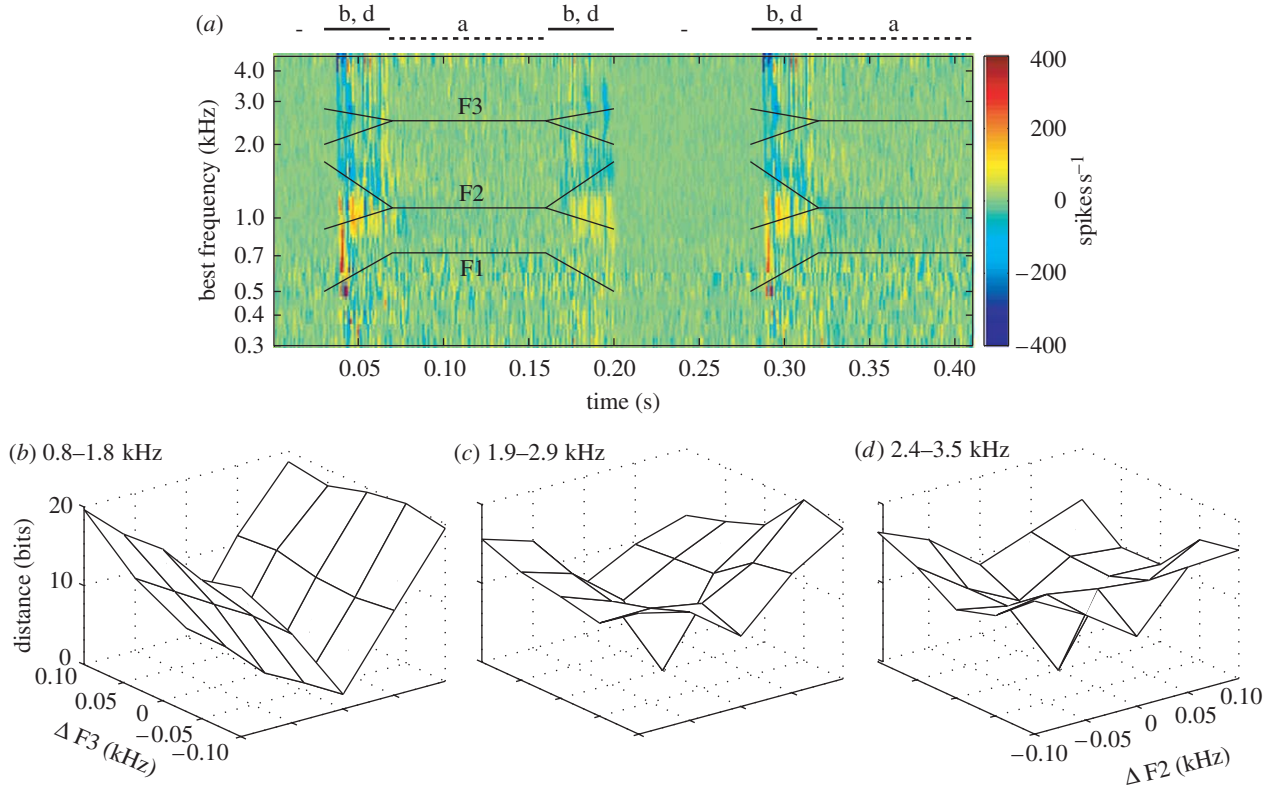

Figure 7. (*a*) Rate differences between responses to /bab ba/ and /dad da/ in a population of AN fibres ($N=137$, containing 78 high SR and 59 LMSR fibres). Fibre BF is plotted on the ordinate and the abscissa is time during the stimulus. The segments of the stimulus are marked at top by the solid (consonant) and dashed (vowel) lines; hyphens mark the silences. The first three formants of both stimuli are shown as black lines. Fibres were gathered in bins by BF (overlapping 0.25 octave bins spaced at 0.0625 octave) and plots of average rate versus time (in 1 ms bins) were constructed for the fibres in each BF bin. The differences in these rates between the two stimuli are plotted on the colour scale identified at right. Positive rate differences mean a higher rate to /bab ba/. The sound level was 70 dB SPL during the vowels. (*b*) Information measure of the dissimilarity of the spike trains of a sub-population of model neurons with BFs in the F2 range (0.8–1.8 kHz). Each point is the dissimilarity between the responses to one of 25 stimuli and the centre stimulus (the distance is zero for the centre stimulus). Dissimilarity is plotted against the frequency differences of F2 (abscissa) and F3 (ordinate) in the stimulus relative to the central reference, measured at 0.06 s. These are labelled at the far left and right only, but are the same in (*b–d*). Only spikes during the interval 0.05–0.07 s were used in the calculation. (*c*,*d*) Same as (*b*) for fibres in higher BF ranges, containing F3 ((*c*), 1.9–2.9 kHz) and above F3 ((*d*), 2.4–3.5 kHz). Contrast the valley shape in (*b*) with the bowl shapes in (*c*,*d*). Adapted with permission from Bandyopadhyay & Young (2004).

lines. The formants are identical during the vowels and diverge during the consonants. /d/ has higher F2 and F3 frequencies than /b/ during the transitions and the F1 frequencies are the same. The colour scale shows the rate differences, warm colours indicating higher rates to /b/. The neural representation shown here is relatively simple (see also Miller & Sachs 1983; Sinex & Geisler 1983). During the vowels and silences, the rates do not differ. During the formant transitions (near 0.05, 0.18 and 0.3 s), rate differences are positive (warm colours) at BFs along the /b/ transitions in F2 and negative (cold colours) at BFs along the /d/ transitions in F2, as expected from the spectra. Rate differences are negative at all BFs near F3 because the /d/ has higher overall energy in this frequency range.

Particularly strong differences are seen in response to the release bursts of the consonants, which occur just after the beginning of the formant transitions at the syllable onsets and last approximately 0.01 s. The time delay is the latency of the neural response. The burst is more high-pass in the /d/ than the /b/ (Blumstein & Stevens 1979) so the rate differences tend to be negative at high BFs and positive at low BFs. The rate differences at the highest BFs (above 3 kHz) are mostly due to the burst.

Discriminability of these CV syllables was quantified by calculating the dissimilarity of the spike trains produced by the two stimuli; the dissimilarity measure is a more general form of the d' measure used in figure 6. If the data have Gaussian distributions, the two measure are proportional (dissimilarity$=d'^2/2\ln 2$). Thus, the interpretation of the dissimilarity measure is the same as for d' (Johnson *et al.* 2001); the reasons for using the dissimilarity measure and the methods for computing it are beyond the scope of this article and are explained by Johnson *et al.* (2001).

As expected from the rate plot, response dissimilarity was large during the formant transitions and zero during the vowels and silences. An unexpected result is that the largest differences, in terms of dissimilarity (and therefore also discriminability of the responses), were seen in high-BF fibres (greater than 3 kHz) during the onset burst. Thus, the stimulus bursts provide substantial information for discriminating stop consonants (Stevens & Blumstein 1978; Smits *et al.* 1996).

The phase-locking analysis (figure 3) suggests that fibres should encode information about formants nearest their BFs. However, the results shown in figure 7*b–d* suggest that this is not necessarily so for high-BF fibres. For this analysis, a stimulus set was synthesized that had F2 and F3 frequency transitions that varied independently over the full range from the transitions of /b/ to the transitions of /d/. Five different values of the frequencies at which the F2 and F3 transitions begin were used and 25 different stimuli with all possible combinations of F2 and F3 transitions were synthesized. The natural /b/ and /d/ lie at two opposite corners of this stimulus set. Responses to the stimulus set were computed using an AN model that has been tuned to produce accurate responses to speech (Bruce *et al.* 2003). Model data were used because it is difficult to obtain the needed amount of data from AN fibres. Differences between the responses to the 24 outlying stimuli and the central stimulus (i.e. the one with the median F2 and F3 frequencies) were computed using the dissimilarity method discussed above, for three BF ranges.

Figure 7*b* shows the result for model fibres with BFs in the range occupied by the F2 transitions. The plot shows the dissimilarity between spike trains as a function of the starting frequency of the F2 (abscissa) and F3 (ordinate) transitions. Distance grows as F2 changes, but not as F3 changes, giving the V-shaped valley shown in the figure. Thus, fibres with BFs near F2 code for F2 only and provide little information about F3. In contrast, fibres with higher BFs give bowl-shaped dissimilarity functions (figure 7*c*,*d*) showing that these fibres convey information about both F2 and F3. This result seems to be inconsistent with the phase-locking analysis of vowels (figure 3) which suggests that fibres with BFs near F3 should respond mainly to F3. It probably reflects the fact that the F2 frequency affects the level of the stimulus near F3, even with fixed F3, owing to the finite bandwidth of the F2 resonance. Thus, even though the neurons with BFs near F3 are responding primarily to F3, they still provide information about both F2 and F3, owing to interactions within the stimulus.

9. FURTHER COMMENTS ON THE REPRESENTATION IN THE AN

The analyses in figures 5–7 suggest that neural codes based on discharge rate are sufficient to represent the first two formants of speech. However, this work considered the relatively unchallenging situation of speech presented in quiet. More difficult situations, such as high sound levels, noisy backgrounds or when more than one person is talking, may require more information than is encoded by rate (Sachs *et al.* 1983; Geisler & Gamble 1989; Palmer 1990; Silkes & Geisler 1991; Keilson *et al.* 1997). Indeed, measures of the discriminability of vowel F2 frequency based on rate show substantial increases in the predicted jnd at signal-to-noise ratios of 3 dB, because the rate change produced by a formant frequency change decreases in the presence of background noise (May *et al.* 1996, 1998)

In all of these studies, information about the vowel remained in the phase-locked responses of fibres, even at signal-to-noise ratios where there was little or no information in rate. Alternative mechanisms for encoding information about speech and other sounds have been suggested that take advantage of the information in phase-locking (Shamma 1985; Deng & Geisler 1987*a*; Carney 1990; Carney *et al.* 2002; Colburn *et al.* 2003). The problem posed by phase-locking is finding a plausible neural mechanism to extract information encoded in this way. One example is a model that detects coincidences between spike trains in AN fibres of different BFs; such a model can produce sharpening of the neural representation over those shown in figure 5 and might also improve performance in noise. Coincidence mechanisms are plausible as a neural mechanism for stimulus representation because they are used in the medial superior olive to extract information about interaural time differences (Goldberg & Brown 1969; Yin & Chan 1990).

The assumption implicit in the work described above, that the neural representation of dynamic speech is a rapid temporal sequence of independent responses to different spectra, is not correct. In fact, there are substantial interactions between the responses to successive segments of speech. Delgutte & Kiang (Delgutte 1980; Delgutte & Kiang 1984*a*) evaluated the effects of a speech sound on the AN response to the following sound. For example, stop consonants like /ba/ have a sudden onset of sound (i.e. increase in sound intensity) where voicing begins; these sound onsets produce brief bursts of high-rate spikes in AN fibres at the time of the onset. These onset bursts can be quite important in neural representations of speech sounds, as for the onset bursts in figure 7. If the same physical sound is preceded by a vocal 'murmur' typical of a nasal consonant, then the overall result sounds like /ma/; in the presence of the murmur, the burst of spikes is significantly reduced or even disappears, even though the vocalic portion of the stimulus is unchanged. The burst disappears owing to short-term adaptation of the AN fibres by their responses to the /m/ (Smith 1977, 1979). The adaptation effect is large enough to change the shape of the rate profile of responses to the vowel. Similar effects were observed when the segment /da/ was preceded by brief sounds that changed the perception of the /da/ to another speech sound (/ada/, /na/, /sha/ or /sa/). Again the major effect was the loss of the burst of spikes produced by the onset of the vowel. Such an onset burst can also signal the rise time of the stimulus, as for the difference between /shoo/ (slow rise) and /chew/ (rapid rise). However, the response bursts are vulnerable to noise masking (Delgutte 1980), so may not provide a robust cue in real-life situations.

10. CHANGES IN THE NEURAL REPRESENTATION OF SPECTRAL SHAPE IN THE COCHLEAR NUCLEUS

The fibres of the AN terminate in the cochlear nucleus, the first auditory centre in the brain. Between the cochlear nucleus and the thalamus, there is a series of complex neural structures making up the brainstem and midbrain auditory circuits (reviewed by Rouiller 1997). The transformations of the neural representation of sound in these systems vary and can be substantial. In the §10–§12, some results on the

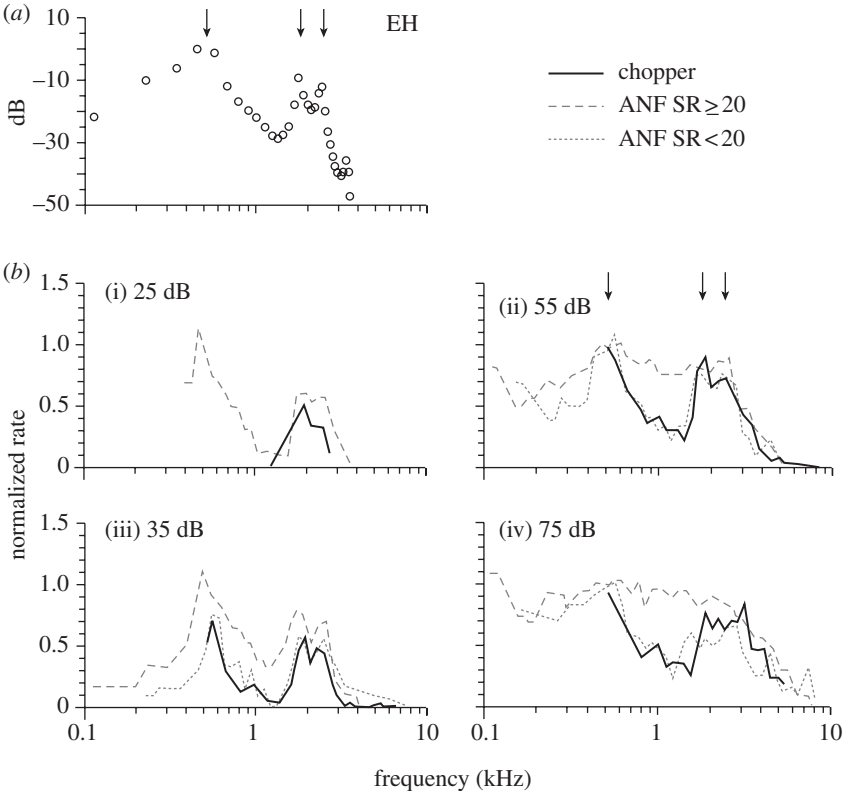

Figure 8. Rate profiles for responses of chopper neurons in the cochlear nucleus of anaesthetized cats to the vowel /ɛ/. The data are actually from the subpopulation of chop-T neurons, but are typical of chopper responses in general. A definition of choppers and a description of their characteristics are provided by Blackburn & Sachs (1989). (*a*) Spectrum of the stimulus, a steady periodic synthetic approximation to /ɛ/ with formants 0.512, 1.792 and 2.432 kHz (arrows) and fundamental frequency 0.112 kHz. (*b*) Rate profiles for a population of chop-T neurons and comparison populations of AN fibres; chopper (solid lines), ANF SR $\geq$20 (dashed lines) and ANF SR $<$20 (dotted lines). Rate is plotted as normalized rate, as in figure 5 and only the log-triangular moving-window averages are shown. AN fibres are divided into LMSR fibres (SR$\leq$20 spikes s^{-1}) and high SR fibres (SR$>$20 spikes s^{-1}). Profiles are shown at four sound levels ((i) 25, (ii) 55, (iii) 35 and (iv) 75 dB), as marked on the plots in dB SPL. Profiles are not plotted at BFs where few neurons were sampled. LMSR AN fibres mostly did not respond at 25 dB SPL, so they are not plotted. AN data adapted with permission from Young & Sachs (1981) and chopper neuron data adapted with permission from Blackburn & Sachs (1990).

representation of speech sounds in central auditory nuclei are discussed, focussing on the differences between the representation in the AN and in the central structure and on the potential importance of those differences. Work on the central neural representation of speech is fragmentary at present, so the results discussed below are examples and are far from a comprehensive view of the subject.

The cochlear nucleus contains between five and ten independent neural subsystems operating in parallel (Rhode & Greenberg 1992; Romand & Avan 1997; Young & Oertel 2003). Each of these receives synaptic inputs from AN fibres of all BFs; in principle, each subsystem constitutes a full neural representation of the sound at the ear. These subsystems differ in terms of synaptic organization, the processing properties of the neurons, and their connections to the rest of the auditory system. Responses to speech have been studied in only a few of these subsystems, mainly the so-called primary-like and chopper neurons. Primary-like neurons relay information important for sound localization to the superior olivary nuclei; for this purpose, it is important to preserve information encoded at high time resolution in the AN (Yin 2002), so these neurons often give responses that are little changed from those of AN fibres. Their responses to speech are generally similar to those described above for AN fibres (Blackburn & Sachs 1990; Winter & Palmer 1990; May *et al.* 1998; Recio & Rhode 2000).

Chopper neurons provide a representation of the spectrum of speech sounds that is improved over that in the AN by being more robust. There are two aspects of the robustness: first, the responses of chopper neurons are less affected by sound level and background noise than are the responses of AN fibres (Blackburn & Sachs 1990; May *et al.* 1998; Recio & Rhode 2000); second, the chopper neurons have a higher sensitivity, or 'gain', for responses to spectral features like the formants (May *et al.* 1996, 1998).

Figure 8 shows a comparison of rate profiles in response to the vowel /ɛ/ from a population of chopper neurons in the cochlear nucleus (heavy solid lines; Blackburn & Sachs 1990) and from AN fibres (dashed and dotted lines; Young & Sachs 1981). The profiles are shown at four sound levels, indicated on the plots. As sound level increases, the profiles of the AN fibres change in that rates increase at all BFs; for the high SR neurons (SR$\geq$20 spikes s^{-1}), the rates saturate at the fibres' maximum discharge rate (a normalized rate near 1) at 55–75 dB. In low and medium SR neurons (SR$<$20 spikes s^{-1}), the rates increase but do not saturate at the sound levels shown. The chopper

neurons' rates also do not saturate and they retain a rate representation that is as good as that of the low SR AN fibres. This result shows that chopper neurons have a mechanism for regulating their dynamic range, for keeping their responses within the dynamic portion of their input/output characteristic.

Except at the lowest sound level (25 dB SPL), which is below threshold for most of the low and medium SR fibres, the results could be explained if choppers receive synaptic inputs only from low and medium SR fibres. However, chopper neurons have low thresholds like high-SR AN fibres and respond to the vowel at low sound levels where low and medium SR fibres do not. Moreover, individual chopper neurons are known to receive synaptic inputs from all SR groups on the basis of cross-correlation evidence (M. B. Sachs & E. D. Young 1988, unpublished data). These characteristics suggest that choppers are able to respond to either high or LMSR fibre populations and that they switch their responses from the former to the latter as sound level increases, called the selective listening hypothesis (Lai *et al.* 1994).

The changes in the representation in the cochlear nucleus are shown in another way in figure 9 (May *et al.* 1996, 1998), where the goal is to analyse the dynamic range of the neurons' responses to the vowel. For this approach, the spectrum of the vowel /ɛ/ was shifted, by changing the sampling rate of the D/A converter used to present the stimulus, to align different features of the vowel with the BF of the neuron under study. Changing the sampling rate shifts the stimulus spectrum along a logarithmic frequency axis without changing its shape. The features were the first three formants (F1, F2 and F3) and four troughs or minima in the spectrum (T0, T1, T3 and T3), defined in figure 9*a*. Examples of the resulting spectra for a neuron with BF 2.1 kHz are shown in the inset of figure 9*a*, for F1, T1 and F2 aligned with BF. At each alignment, the neuron's response rate was measured at several sound levels. The discharge rates for one example AN fibre are plotted in figure 9*a* by the solid lines and open circles; this plot shows discharge rate versus the frequency of the feature, aligned with the spectrum at the nominal sampling rate (i.e. the one used in previous figures).

Figure 9*b* shows the average driven discharge rates (rate − spontaneous rate) of two populations of AN fibres (separated by SR) plotted against the sound level of the feature that is aligned with BF, i.e. of the formant or trough harmonic. In each case, data are shown for different values of the overall sound level of the stimulus, plotted with different symbols. The data points taken at one overall sound level were fitted by a straight line, whose slope is a measure of the gain of the neuron, in spikes/(s dB)$^{-1}$, at the corresponding sound level. The gain is a measure of the expected quality of a rate profile constructed with these neurons; the larger the gain, the larger the peak height at the F2 frequency.

For the high-SR AN fibres (HSR (*b*(i))), the rates resemble a rate-versus-level function for a BF tone. In particular, the slope of the rate function (or of the lines fitting the points) decreases at high sound levels as the fibre's rate response saturates. In contrast, there is little saturation at levels between 43 (filled squares) and 93 dB SPL (down-pointing triangles) for the LMSR

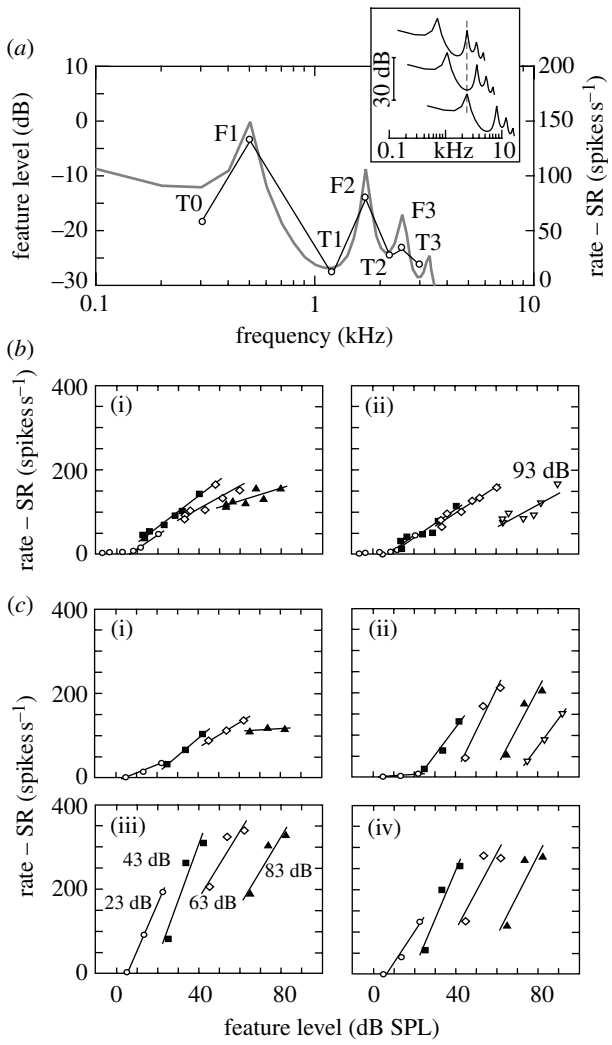

Figure 9. (*a*) Spectrum of the synthetic /ɛ/ (grey solid line, left ordinate) and driven discharge rates (black lines with open symbols, right ordinate) of an AN fibre when the vowel spectrum was shifted along the log frequency axis by changing the sampling rate so as to align various spectral features with BF. The vertical alignment of the left and right ordinates is arbitrary. The inset shows the spectra of three such shifted vowels that align F2, T1 and F1 (in order top to bottom) with the 2.1 kHz BF of the fibre, at the dashed line. (*b*) Driven discharge rate (rate–SR) plotted versus the sound level of the stimulus feature aligned on BF; data are averages for high (i) and low+medium (ii) SR populations of AN fibres. (*c*) The same plots for four groups of cochlear nucleus neurons: (i) high SR primary-like, (ii) low SR primary-like, (iii) and (iv) chopper. These neuron types are defined elsewhere (Blackburn & Sachs 1989) but the differences are not important for this discussion. In both (*b*,*c*), the shifted vowels were presented at three or four overall sound levels, shown with different symbols. The sound levels are defined in (*b*(ii)) and (*c*(iii)) as dB SPL. The feature level on the abscissa is the sum of the overall sound level of the vowel and the relative level of the feature and is the SPL of the stimulus harmonic that is located at BF. For the AN data, the vowel was shifted to seven positions relative to BF, as in (*a*); for the cochlear nucleus neurons, only three positions were used (F1, T1 and F2). Adapted with permission from May *et al.* (1997)).

fibres (*b*(ii)). Moreover, there is substantial dynamic range adjustment as the overall level of the stimulus changes, seen by comparing the rates in response to vowels at two different overall sound levels at a fixed

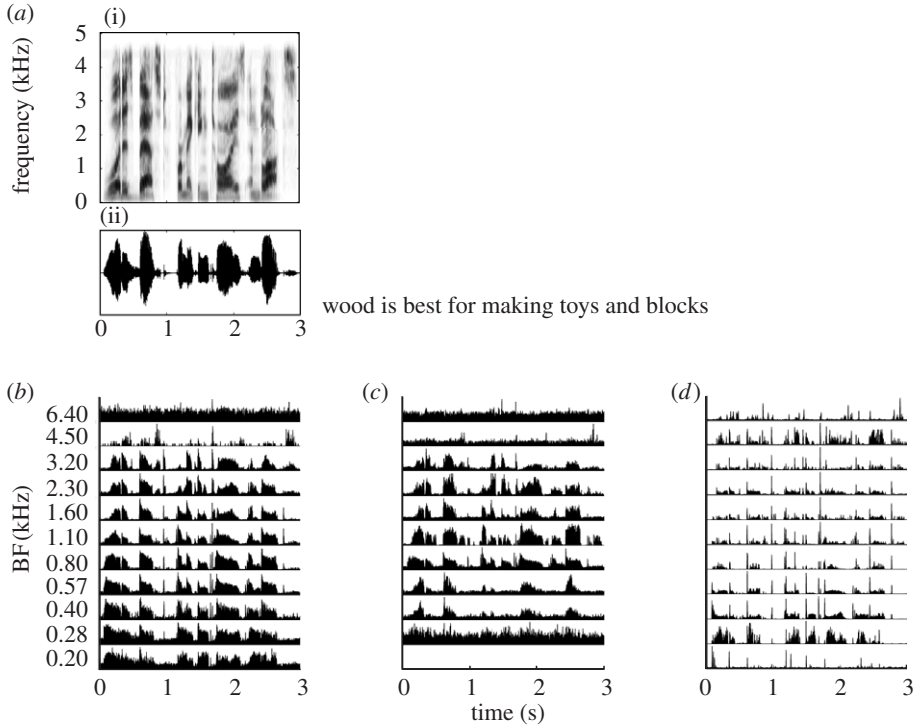

Figure 10. (*a*(i)) Spectrogram and (*a*(ii)) oscillogram of the sentence 'wood is best for making toys and blocks'. (*b*)–(*d*). PST histograms of responses of populations of AN fibres (*b*), neurons from cochlear nucleus (*c*) and neurons from inferior colliculus (*d*) to this sentence. The responses of the neurons were recorded in anaesthetized cats. The PST histograms are the averages of responses of populations of neurons gathered into 0.5 octave bands, with centre frequencies indicated at left. Adapted with permission from Delgutte *et al.* (1998).

feature level. For example, for the LMSR fibres, the rate in response to a feature level of, say, 60 dB SPL is considerably smaller for the 93 dB SPL overall stimulus level (where it is a response with T1 at BF) than for the 63 dB SPL stimulus level (with F1 at BF). This behaviour is also observed for the HSR AN fibres, but it is a smaller effect.

Figure 9*c* shows the same analysis for four populations of cochlear nucleus neurons. The primary-like neurons (top row) show behaviour similar to the AN fibres in figure 9*b*. The chopper neurons (bottom row) behave like the LMSR fibres; again, the slopes of the lines do not decrease much as the overall stimulus level increases and the dynamic range adjustment, as defined above, is clear.

Dynamic range adjustment means that the discharge rate with a particular feature at BF depends on whether the feature is located at a peak of the stimulus spectrum or a minimum. The lowest rates are for troughs aligned with BF and the highest rates are for spectral peaks aligned with BF. Presumably, this dynamic range adjustment is a consequence of suppression or inhibition of the response to the weaker components of the stimulus by the stronger components.

There is an important systematic difference between the chopper and LMSR primary-like neurons versus the AN fibres in that the slopes of the lines are significantly higher for the cochlear nucleus neurons, by up to a factor of two (May *et al.* 1998). The increased gain combines with the stability of dynamic range adjustment in the chopper and LMSR primary-like neurons to provide an improved spectral representation. As expected from these data, the jnd for F2 frequency calculated for the chopper neurons is less than or equal to that for the AN and is stable across sound level (May *et al.* 1996). The difference extends to the representation in the presence of background noise. The slopes of lines like those in figure 9*b* decrease in the presence of background noise, but more so for AN fibres than for chopper neurons; the slope decrease between a vowel in quiet and a vowel in background noise at a signal-to-noise ratio of 3 dB is about twice as large in AN fibres as in chopper neurons.

11. THE TEMPORAL ENVELOPE OF SPEECH IN THE AN AND INFERIOR COLLICULUS

The speech signal is strongly modulated in time. An example is shown in figure 10*a* (Delgutte *et al.* 1998). The spectrogram (top) and an oscillogram (bottom) of the sentence 'wood is best for making toys and blocks' are shown. Associated with the syllables of this sentence are strong modulations of its envelope, seen as changes in the amplitude of the oscillogram at frequencies of a few hertz. Generally, the vowels have the most energy, shown by the largest amplitudes in the oscillogram and the corresponding dark parts of the spectrogram. Silences associated with the stop consonants are seen as near-zero amplitude parts of the oscillogram and light parts of the spectrogram.

The emphasis of this paper has been on the neural representation of the frequency content of speech; however, important information is encoded in the temporal envelope as well (Van Tasell *et al.* 1987; Rosen 1992; Shannon *et al.* 1995). Examples have been mentioned earlier, such as the stimulus rise time differences between the consonants /ch/ and /sh/ and the representation of rise time in terms of transient or

onset bursts of spikes in AN fibres (Delgutte & Kiang 1984*a*). The best-studied temporal contrast in speech is the voice-onset time (VOT), which is important in differentiating voiced from voiceless stop consonants (e.g. /b/ from /p/).

Figure 10 shows data from Delgutte *et al.* (1998) on the neural representation of the temporal envelope of a sentence, without reference to a particular speech sound. The histograms in figure 10*b–d* show respectively the average responses of populations of AN fibres, neurons in the cochlear nucleus and neurons in the inferior colliculus to the sentence in figure 10*a*. The neurons were grouped by BF into 0.5 octave bins and the PST histograms of the neurons falling into each bin were averaged. The histograms shown are the averages, identified at left by the centre frequency of the analysis bin.

The AN responses (figure 10*b*) correspond generally to the envelope of the stimulus, in that rates are high during segments of the stimulus with significant energy. Of course, the rates are modified by the frequency spectrum of the different parts of the stimulus. The analysis was not designed to show the representation of the stimulus spectrum, but the effects of spectrum can be seen, for example, at the beginning of the stimulus, where the /w/ has mainly low-frequency energy and the rate responses are higher in low-BF fibres. The responses to transient parts of the stimulus, especially onsets of syllables or bursts associated with stop consonants, are evident as sharp peaks in rate that extend broadly across frequency. For example, the responses to the /t/s are clear just before 1 s and at approximately 1.7 s. Even so, there are substantial AN responses, in the form of maintained rate elevations, to the vowels and the steady consonants (/w/, /s/, /m/, etc.). The responses in the cochlear nucleus (figure 10*c*) are qualitatively similar, although they appear to be more differentiated across frequency bands.

The data in figure 10*d* show the responses of neurons in the inferior colliculus. The colliculus is a structure unique to the auditory system, in that other sensory systems do not have an equivalent. An excellent and comprehensive review of the colliculus is available (Winer & Schreiner 2005). Its neurons collect axons from a number of sources in the brainstem auditory nuclei, including the cochlear nucleus and the superior olivary complex, as well as descending axons from thalamus and cortex (Oliver & Huerta 1992; Rouiller 1997). Its outputs travel to the auditory part of the thalamus, the medial geniculate, which projects in turn to the auditory cortex. Although the functional properties of its neurons have been studied extensively (reviewed by Irvine (1992), Ehret (1997) and several papers in Winer & Schreiner (2005)), an understanding of the mechanisms for processing of complex stimuli in the colliculus has not been achieved.

The responses in the inferior colliculus (figure 10*d*) are more transient than those at lower levels. The strongest responses are to the onsets and bursts in the speech and steady responses to the vowels have relatively low discharge rates. These data suggest that the central representation of speech may emphasize transients over steady-state responses, which should have the effect of amplifying the representation of consonants relative to vowels.

Delgutte *et al.* (1998) used a simplified neural model to represent the transformations shown in figure 10. The model consisted of a bandpass filter to model peripheral tuning curves, followed by a rectifier to model compression and to extract the envelope from the responses, followed by a second filter to model the neuron's modulation sensitivity. Neural responses to modulation are generally characterized using sinusoidal amplitude modulation of a BF tone or broadband noise (reviewed by Joris *et al.* 2004). The sensitivity of the neuron for modulation is computed as the modulation gain, the neural response divided by the stimulus modulation amplitude; when the gain is plotted against modulation frequency, the result is the modulation transfer function (MTF).

MTFs were measured for the neurons used for figure 10. The MTFs of the three different groups of neurons were bandpass functions that were similar in terms of the magnitude of the gain versus modulation frequency, except that the inferior colliculus neurons were less sensitive at higher modulation frequencies (above approx. 100 Hz; Joris *et al.* 2004). More important, there was a difference in the phase characteristics of the MTFs; this difference caused the inferior colliculus neurons to give transient responses to a steady stimulus as opposed to tonic responses in the AN and the cochlear nucleus. In effect, the inferior colliculus neurons behaved as if their response consisted of an excitatory component followed at short latency by an inhibitory component. In the AN and cochlear nucleus, the inhibitory component was not seen or was weaker. Although this model did not predict all the details of the responses, it was qualitatively consistent with the data shown in figure 10 in that it produced phasic responses in colliculus neurons.

The VOT is the delay of the onset of voicing from the release time of a stop consonant. For the voiced stops (/b/, /d/ and /g/) voicing begins as soon as the closure of the vocal tract is released. The stops analysed in figure 7 are of this type where the VOT is essentially zero. For the unvoiced stops (/p/, /t/ and /k/), there may be a 40–80 ms delay before voicing begins. This pause is signalled in the acoustic signal as a lack of low-frequency energy in the signal prior to the onset of voicing; thus, there is little energy at F1 prior to the onset of voicing, while F2 and F3 contain energy produced by the noise associated with release of the stop.

The representation of VOT has been analysed by Sinex and colleagues (Sinex & McDonald 1988; Sinex *et al.* 1991; Sinex 1993; Sinex & Narayan 1994; Chen *et al.* 1996). As expected from the description of the stimulus above, it is mainly low-BF neurons (less than 1 kHz) that behave differently between voiced and voiceless stops and the difference is that low-BF neurons show a pause in discharge during the voiceless period and an onset of spiking when voicing begins. The duration of the pause in discharge corresponds well to the VOT (Sinex *et al.* 1991) and serves as a neural correlate of the VOT that could be used by the brain to discriminate voiced and voiceless stops.

The representation of VOT seems to change little between the AN and the inferior colliculus. At both levels, there is a pause in spiking whose duration

corresponds to the VOT. Conversion of the duration of the pause to some other form of neural code, like a discharge rate proportional to the pause, apparently does not occur (Sinex & Chen 2000) even though there are duration-tuned neurons in the inferior colliculus in bats (Pinheiro *et al.* 1991; Ehrlich *et al.* 1997; Fuzessery & Hall 1999) and other mammals (Chen 1998; Brand *et al.* 2000).

12. CORTICAL REPRESENTATION OF TEMPORAL AND SPECTRAL PARAMETERS

Stimulus representations in the auditory cortex have many aspects that differentiate them from the peripheral representations summarized above (e.g. Kaas *et al.* 1999; Schreiner *et al.* 2000; Eggermont 2001; Nelken 2004; Metherate *et al.* 2005). The primary auditory cortex is organized tonotopically, so that there is an orderly arrangement of neurons by BF, with rows of neurons tuned to particular frequencies arranged side by side in order from low to high frequency. Perpendicular to the tonotopic axis (within a row), the neurons are all tuned roughly to the same frequency, but can differ in a number of other response parameters; these include binaural sensitivity, width of tuning, strength and arrangement of inhibitory inputs, and dynamic range properties. Thus, the representation of stimuli in auditory cortex is guaranteed to be more diverse and multifaceted than the straightforward representation in peripheral neurons.

Even for tones, the organization suggested by the tonotopic map can be misleading. The tonotopic map is defined by the BFs of neurons at low sound levels, within a few dB of threshold. At higher sound levels, inhibitory inputs can reshape the responses so that the maximum response to a tone of a particular frequency can lie outside its own tonotopic location and neurons at the tonotopic location can be inhibited (e.g. Schreiner 1998, fig. 2).

There are a number of aspects of cortical physiology that make addressing the question of the cortical 'representation' of speech difficult, or at least substantially different from the more peripheral representations. Three such aspects are discussed below.

First, cortical neurons are often specifically selective for sounds that are behaviourally important for the animal. In marmosets, for example, the responses of cortical neurons to the species' vocalizations are strongly different if the sounds are reversed in time (Wang & Kadia 2001); that this result is not a consequence of general auditory processing follows from the lack of an effect of time reversal when the same sounds are presented to neurons in the cat auditory cortex. Other well-studied examples of specificity for species-specific sounds are the songbirds' auditory cortex equivalent (field L), where neurons are selective for song relative to similar artificial sounds (Grace *et al.* 2003; Cousillas *et al.* 2005), and the songbirds' auditory motor song nuclei, where neurons are highly selective among songs according to singer (Margoliash 1986; Doupe & Konishi 1991). These examples make the point that cortical structures can be organized to respond specifically to behaviourally meaningful categories of sounds rather than to the general spectrotemporal properties of sounds; that is to respond to sounds as objects, not just sounds (Nelken *et al.* 2003).

Second, although neurons in primary auditory cortex have BFs and receptive fields that can be defined with methods like tuning curves or reverse correlation, the receptive fields are not fixed properties of the neurons. For example, in ferrets trained to perform a behavioural task for a food reward, the BF of a neuron can shift over a significant frequency range until it is centred on the stimulus frequencies important for the task (Fritz *et al.* 2003); the shift may be temporary and reverse when the behavioural task is removed or it may be long-lasting. A second example is provided by the phenomenon of oddball responses, in which the response of a neuron to a particular stimulus depends on its probability of occurring: common sounds evoke weaker responses than rare sounds (Ulanovsky *et al.* 2003). Thus, it is misleading to think of the characteristics of cortical neurons as fixed; in fact, they may be adjusted to immediate behavioural contingencies and ongoing sound backgrounds in order to optimize the processing of sound in some way.

Third, neurons in auditory cortex may respond to sounds in ways that cannot be accurately modelled by straightforward concepts of receptive fields. At the most basic level, this comment refers to the inadequacy of the spectrotemporal receptive field (STRF) model. Such models are fitted to the responses of neurons to a suitable stimulus set, typically broadband noise or some other broadband stimulus (Aertsen & Johannesma 1981; Eggermont 1993; Klein *et al.* 2000; Theunissen *et al.* 2001). The method is similar to reverse correlation discussed in connection with figure 1*b*, except that a measure of the power spectrum of the stimulus preceding spikes is averaged. The STRF model is tested by using it to predict the responses of the neuron to another stimulus set; typically, such models account for less than half the variance in the responses of cortical neurons (Yeshurun *et al.* 1989; Versnel & Shamma 1998; Sen *et al.* 2001; Machens *et al.* 2004).

A more subtle indication of the inadequacy of STRF models is that a neuron presented with a complex acoustic signal may not respond to the most intense frequency components near the neuron's BF. For example, when presented with a bird chirp that is accompanied by echoes and background noise, neurons in cat cortex may respond to the whole stimulus differently than to the chirp alone, even when the chirp is at a frequency near BF and has a sound level well above that of the other components (Bar-Yosef *et al.* 2002). This result is consistent with the conclusion drawn above that the responses of cortical neurons are a step removed from the immediate spectrotemporal aspects of the stimulus (Nelken 2004).

An example of the responses of a population of neurons in the auditory cortex of the cat to the syllables /be/ and /pe/ is shown in figure 11 (Schreiner 1998; Wong & Schreiner 2003). The stimuli differ in VOT; their spectra are shown in figure 11*a*. The formants, marked in figure 11*a*(ii), are reasonably steady in time. The VOT is approximately 70 ms for the /p/ and much shorter for the /b/. In the spectrogram, the release burst of the /p/ is just visible as a broadband signal near time zero. A similar burst is present for the /b/, where it is

accompanied by the large low-frequency (less than 1 kHz) energy in the voicing. Figure 11*b* shows neurograms of the responses of a population of cortical neurons to the sounds. Neurons are displayed along the ordinate according to their BFs; each horizontal line shows the firing rate of neurons in a particular BF bin plotted on the same time axis as the spectrograms in figure 11*a*. In this anaesthetized preparation, the neurons respond mainly to the onset of the sound at the release of the /p/ or /b/ (just after time zero) and to the onset of the voicing (at 70 ms) in the case of /pe/. For /be/, there is no change in the voicing to produce a second response. Looking vertically down the ordinate in figure 11*b* shows the tonotopic distribution of responses. Although there are some changes in latency (time of response) with BF, there are not obvious rate peaks in these data that correspond to the formants of the signals.

The responses to stop-consonant syllables shown in figure 11*b* are similar to those observed in other studies, done using anaesthetized cats (Eggermont 1995) and awake monkeys (Steinschneider *et al.* 1995, 2003). With sufficiently long VOTs, two peaks of response are seen, as for /pe/ in figure 11*b*, one corresponding to the release of the stop consonant and the second to the onset of voicing. As the VOT is made shorter, the second peak disappears. Although there are other components of the cortical responses to these syllables, especially in awake monkeys, the existing evidence again points towards the pause in discharge between two onset response peaks as the representation of VOT.

Figure 11*c* shows the distribution of activity across the region of cortex from which recordings were made by Wong & Schreiner (2003). The four plots in figure 11*c* show maps of the response strength of neurons as a function of location. The estimates of response are based on single and multi-unit recordings from 88 recording sites distributed uniformly across the map. The colour scale shows the number of spikes during two time windows. The first window (1–50 ms; figure 11*c*(i)(iii) contains the burst of response to the consonant release and the second window (51–100 ms, figure 11*c*(ii)(iv) contains the onset of voicing. The tonotopic map is marked on the cortical surface by the black lines, which show the approximate locations of neurons tuned to the first three formants of the stimuli.

Consistent with figure 11*b*, there is no response to /be/ in the second time interval (figure 11*c*(ii)). Looking at the other three maps, one sees patchy distributions of responses which do not necessarily peak at BFs corresponding to the formant frequencies. Although the responses to the two stimuli are clearly different, there is no simple correspondence between the stimulus spectrum and the tonotopic map. It is particularly convincing to look along an isofrequency contour corresponding to one of the formants. Substantial differences in response are observed in all cases, except for F1 in the lower right map. The differences in responses within an isofrequency sheet are as large as the differences across the tonotopic map.

These experiments have not been repeated elsewhere, although similar results have been shown for responses to species-specific communication sounds in marmoset cortex (Wang *et al.* 1995; Wang 2000). In the marmoset, neurons give strong burst responses to successive syllables of the stimulus, analogous to the two-peak responses to the VOT stimuli. There is a tonotopic representation of call spectrum in the marmoset, which varies according to the degree of selectivity of the neurons for the calls. However, maps of responses like figure 11*c* were not obtained.

The experiments summarized in figure 11 illustrate the problems of studying speech representation in cortex. In thinking about the data in figure 11, it is worthwhile to consider the following hypothesis about the organization of cortical responses beyond the tonotopic gradient. Chi *et al.* (2005) have proposed a multiparameter representation of speech modelled on the properties of cortical STRFs, which measure both frequency tuning and temporal responses of neurons. In the frequency domain, the width of frequency tuning and the arrangement of inhibitory versus excitatory regions change the selectivity of the neuron for broad versus narrow resonances in the stimulus spectrum. This aspect is called 'scale'. Each neuron is a bandpass filter for stimulus scale, and is most sensitive to a particular scale; the tuning for scale can be derived from the magnitude of the Fourier transform of the STRF along the frequency axis. Low-scale neurons would respond best to speech sounds with broad, widely spaced formants and high-scale neurons would prefer narrow, closely spaced formants. A second parameter is 'rate' which measures the selectivity of the neuron for FM sweeps. In the model, neurons are sensitive to a range of sweep rates, both upward and downward. Rate is important in representing formant transitions, for example. A speech sound can be completely represented in terms of the scale and the rate as a function of frequency, meaning that the stimulus can be reconstructed accurately from the description in terms of scale and rate.

Now consider what the representation of a speech sound would be in a cortex that is organized in this way. Along an isofrequency sheet, there would be neurons sensitive to different rate and scale. A sound with a formant at a particular frequency might or might not activate a neuron with a BF at the frequency of the formant, depending on the relative scales and rates of the stimulus and the neuron. Such a representation could be patchy as in figure 11*c*. This is not meant to suggest that the cortex is actually organized in this way; rather, this hypothesis serves as an example of a representation that could not be easily analysed, from data like figure 11*c*, without some foreknowledge of the principles upon which it is based.

13. HOW IS SPEECH REPRESENTED BY NEURONS?

The nature of the neural representation of speech is clearest in the case of AN fibres, where the spectro-temporal features of the stimulus are isomorphically represented by neural activity. The frequency content of sounds is represented by the distribution of activity across BFs in the population of AN fibres, as displayed in a rate profile. The temporal envelope of connected speech and specific temporal features of the speech signal like the VOT are represented directly by temporal modulation of the neural response.

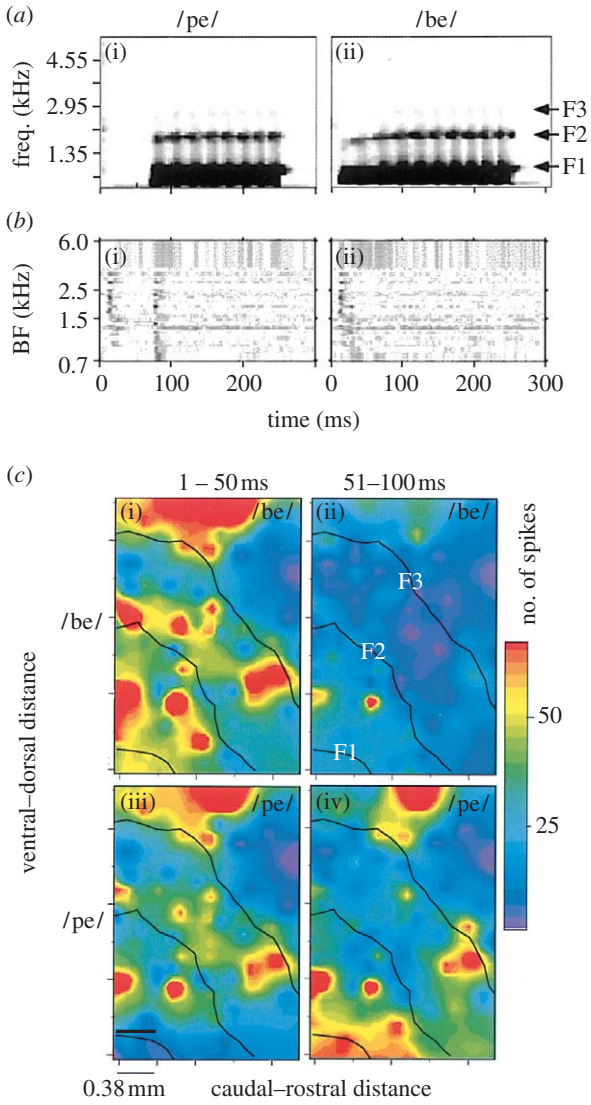

Figure 11. Population responses of neurons in anaesthetized cat cortex to two CV syllables. (*a*) Spectrograms of the syllables (i) /pe/ and (ii) /be/ used as stimuli. The formants are roughly constant (0.6, 1.7 and 2.5 kHz) and are marked at (ii). The VOTs are 70 ms for /pe/ and near-zero for /be/. (*b*) Neurograms showing discharge rate as a function of time (on the abscissa) for neurons of different BFs (on the ordinate). The darkness of a pixel corresponds to firing rate on a normalized scale from 0 (white) to half the rate at the sound level where the BF tone rate versus level function begins to saturate (black). There are few neurons for BFs above 3.6 kHz, so the last few BF bins are combined. (*c*) Plots of discharge rate (colour scale) versus the location of the recording electrode across the surface of auditory cortex. Data are interpolated from responses at 88 recording locations within the map. Maps are shown for (i) and (ii) /be/, and (iii) and (iv) /pe/ over two time intervals: (i) and (iii) (1–50 ms), (ii) and (iv) (51–100 ms). The tonotopic map across this piece of cortex is indicated by showing the approximate isofrequency lines for the first three formants, labelled in white in (ii). The firing rate is given as the number of spikes in the analysis interval during 20 repetitions of the stimulus. Data in (*b*) and (*c*) are from different brains, but the results are qualitatively the same; in both cases, the stimulus level was 55 dB SPL. Adapted with permission from Schreiner (1998) and Wong & Schreiner (2003).

Less is known about the nature of the representation of speech in central neurons. Owing to the large number of different groups of auditory neurons in the brain, only fragmentary results on a few of the potentially important transformations are available. For studies in a non-human animal, questions about the representation of speech have to be translated into general questions about the neural representation of complex auditory stimuli; that is, there may be specializations for processing speech in the human brain that are not present in animal models and those specializations could be present at levels below the cortex. In the bat, for example, there are specializations for biosonar processing in the inferior colliculus and higher centres (O'Neill *et al.* 1989; Mittmann & Wenstrup 1995).

The changes in the neural representation of speech described for central auditory neurons in this paper are mainly improvements in the robustness of the representation, as for rate profiles in the cochlear nucleus, and emphasis of some aspects of the response over others, as in the apparent amplification of responses to stimulus transients in the inferior colliculus. These are aspects of generic processing of auditory spectro-temporal representations which might be appropriate for almost any natural sound. Further investigation of the computational mechanisms by which these changes are achieved will be useful both for the general problem of how the brain represents natural auditory scenes and for the specific problem of how lower parts of the auditory system process speech.

The issue raised at the end of the discussion of the auditory cortex is a particularly important challenge. The assumption in most work on the representation of complex stimuli in the auditory system is that the representation is tonotopic and that there is some additional analysis like the scale and rate dimensions discussed above. However, there is no widely accepted theory of what those dimensions should be, much less how they should be represented in neurons. This remains the central problem of research on the neurophysiology of speech and natural stimuli.

Preparation of this paper was supported by NIH grant DC00109. Thanks to Ian Bruce, Michael Heinz, Sharba Bandyopadhyay, Brian Moore and two anonymous reviewers for their comments on a previous version.

APPENDIX A. THE NATURE OF THE DATA

Studies of the neural representation of speech are based on microelectrode recordings of the activity of single neurons or small groups of neurons in anaesthetized animals. A microelectrode is placed into the neural structure of interest and the activity of the neurons near the electrode tip is recorded as speech-like stimuli are presented in the ear. The experiments are done successively on as many neurons as possible, presenting the same stimulus to each. Thus, the experimenter builds up an estimate of the responses to the speech sound in a population of neurons (Pfeiffer & Kim 1975; Sachs & Young 1979). Usually, the electrode is placed so as to isolate the activity of one neuron at a time. Most of the results described here were obtained in this way. In some cases, however, the activity of small groups of unseparated neurons is recorded, so-called multiunit recording, represented here by the data in figure 11.

Many aspects of the activity of neurons can be recorded experimentally. It is beyond the scope of this

paper to describe the physiology of neurons, which can be found in texts devoted to that subject (e.g. Kandel *et al.* 2000; Shepherd 2004). An introduction to the specialized neurophysiology of the auditory system can be found in Pickles (1988). Studies of the neural representation of speech are based on recordings of spike trains, meaning the trains of action potentials produced by the neuron under study. Conventionally, it is assumed that the information in spike trains is encoded entirely in the times of occurrence of action potentials. Thus, it is sufficient, in studying neural representations, just to record and analyse spike times (Rieke *et al.* 1997). Neural responses are the changes in the number or the temporal arrangement of spikes produced by the stimulus.

Before analysing the responses of a neuron, it is necessary to characterize the neuron in terms of known properties that may be expected to influence its responses to the stimulus. In the simplest case of AN fibres, the neurons differ in terms of the frequencies to which they are sensitive, their frequency tuning (figure 1; Kiang *et al.* 1965; Liberman 1978), and in terms of the range of sound intensities to which they respond, their thresholds and dynamic ranges (Sachs & Abbas 1974; Yates *et al.* 1990). The importance of these characterizations is evident from the results shown in figures 3–6.

In the central nervous system, the analysis problem is more complicated because there may be several different groups of neurons operating in parallel. Each group is made up of cells with different anatomical organizations, different electrophysiological properties and different connections to the rest of the auditory system. In principle, each group of neurons forms an independent representation, with information about various aspects of the stimulus represented differentially in different groups. Characterization of neurons in this case must be more extensive, extending beyond frequency tuning and dynamic range to include an estimate of the type of neuron from which a recording is being made and perhaps other properties. The matter of neural typing is best understood for the cochlear nucleus and is reviewed elsewhere (Rhode & Greenberg 1992; Romand & Avan 1997; Young & Oertel 2003).

An additional complication, for neurons in the central nervous system, is the effect of anaesthesia. The studies described above were almost all done in anaesthetized animals, but the effects of anaesthesia on response properties of central neurons are significant (e.g. Kuwada *et al.* 1989; Zurita *et al.* 1994; Ramachandran *et al.* 1999; Gaese & Ostwald 2001; Anderson & Young 2004). The major effects of anaesthetics are a lowering of spontaneous discharge rate, an increase in threshold and a loss of inhibitory responses. The loss of inhibition probably has particularly important effect on responses to complex stimuli, where inhibitory interactions can predominate. Thus, the presence of anaesthesia is a source of uncertainty in interpreting studies of the neural representation of speech in central auditory nuclei (but see May *et al.* (1998) for a counterexample in the cochlear nucleus).

REFERENCES

Aertsen, A. M. H. J. & Johannesma, P. I. M. 1981 The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol. Cybern.* **42**, 133–143. (doi:10.1007/BF00336731)

Anderson, M. J. & Young, E. D. 2004 Isoflurane/N_2O anesthesia suppresses narrowband but not wideband inhibition in dorsal cochlear nucleus. *Hear. Res.* **188**, 29–41. (doi:10.1016/S0378-5955(03)00348-4)

Bandyopadhyay, S. & Young, E. D. 2004 Discrimination of voiced stop consonants based on auditory-nerve discharges. *J. Neurosci.* **24**, 531–541. (doi:10.1523/JNEUROSCI.4234-03.2004)

Bar-Yosef, O., Rotman, Y. & Nelken, I. 2002 Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. *J. Neurosci.* **22**, 8619–8632.

Blackburn, C. C. & Sachs, M. B. 1989 Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. *J. Neurophysiol.* **62**, 1303–1329.

Blackburn, C. C. & Sachs, M. B. 1990 The representations of the steady-state vowel sound /ɛ/ in the discharge patterns of cat anteroventral cochlear nucleus neurons. *J. Neurophysiol.* **63**, 1191–1212.

Blumstein, S. E. & Stevens, K. N. 1979 Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *J. Acoust. Soc. Am.* **66**, 1001–1017. (doi:10.1121/1.383319)

Brand, A., Urban, A. & Grothe, B. 2000 Duration tuning in the mouse auditory midbrain. *J. Neurophysiol.* **84**, 1790–1799.

Bruce, I. C., Sachs, M. B. & Young, E. D. 2003 An auditory–periphery model of the effects of acoustic trauma on auditory nerve responses. *J. Acoust. Soc. Am.* **113**, 369–388. (doi:10.1121/1.1519544)

Carney, L. H. 1990 Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents. *J. Neurophysiol.* **64**, 437–456.

Carney, L. H. 1993 A model for the responses of low-frequency auditory-nerve fibers in cat. *J. Acoust. Soc. Am.* **93**, 401–417. (doi:10.1121/1.405620)

Carney, L. H. & Geisler, C. D. 1986 A temporal analysis of auditory-nerve fiber responses to spoken stop consonant–vowel syllables. *J. Acoust. Soc. Am.* **79**, 1896–1914. (doi:10.1121/1.393197)

Carney, L. H. & Yin, T. C. T. 1988 Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model. *J. Neurophysiol.* **60**, 1653–1677.

Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H. & Colburn, H. S. 2002 Auditory phase opponency: a temporal model for masked detection at low frequencies. *Acustica-Acta Acustica* **88**, 334–347.

Chen, G.-D. 1998 Effects of stimulus duration on responses of neurons in chinchilla inferior colliculus. *Hear. Res.* **122**, 142–150. (doi:10.1016/S0378-5955(98)00103-8)

Chen, G.-D., Nuding, S. C., Narayan, S. S. & Sinex, D. G. 1996 Responses of single neurons in the chinchilla inferior colliculus to consonant–vowel syllables differing in voice onset time. *Aud. Neurosci.* **3**, 179–198.

Chi, T., Ru, P. & Shamma, S. A. 2005 Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.* **118**, 887–906. (doi:10.1121/1.1945807)

Colburn, H. S., Carney, L. H. & Heinz, M. G. 2003 Quantifying the information in auditory-nerve responses for level discrimination. *J. Assoc. Res. Otolaryngol.* **4**, 294–311. (doi:10.1007/s10162-002-1090-6)

Conley, R. A. & Keilson, S. E. 1995 Rate representation and discriminability of second formant frequencies for /ɛ/-like steady-state vowels in cat auditory nerve. *J. Acoust. Soc. Am.* **98**, 3223–3234. (doi:10.1121/1.413812)

Cousillas, H., Leppelsack, H. J., Leppelsack, E., Richard, J. P., Mathelier, M. & Hausberger, M. 2005 Functional organization of the forebrain auditory centres of the European starling: a study based on natural sounds. *Hear. Res.* **207**, 10–21. (doi:10.1016/j.heares.2005.01.008)

de Boer, E. & de Jongh, H. R. 1978 On cochlear encoding: potentialities and limitations of the reverse-correlation technique. *J. Acoust. Soc. Am.* **63**, 115–135. (doi:10.1121/1.381704)

Delgutte, B. 1980 Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *J. Acoust. Soc. Am.* **68**, 843–857. (doi:10.1121/1.384824)

Delgutte, B. 1984 Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds. *J. Acoust. Soc. Am.* **75**, 879–886. (doi:10.1121/1.390597)

Delgutte, B. 1990 Two-tone rate suppression in auditory-nerve fibers: dependence on suppressor frequency and level. *Hear. Res.* **49**, 225–246. (doi:10.1016/0378-5955(90)90106-Y)

Delgutte, B. & Kiang, N. Y. S. 1984*a* Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *J. Acoust. Soc. Am.* **75**, 897–907. (doi:10.1121/1.390599)

Delgutte, B. & Kiang, N. Y. S. 1984*b* Speech coding in the auditory nerve: III. Voiceless fricative consonants. *J. Acoust. Soc. Am.* **75**, 887–896. (doi:10.1121/1.390598)

Delgutte, B. & Kiang, N. Y. S. 1984*c* Speech coding in the auditory nerve: I. Vowel-like sounds. *J. Acoust. Soc. Am.* **75**, 866–878. (doi:10.1121/1.390596)

Delgutte, B., Hammond, B. M. & Cariani, P. A. 1998 Neural coding of the temporal envelope of speech: relation to modulation transfer functions. In *Psychophysical and physiological advances in hearing* (eds A. R. Palmer, A. Rees, A. Q. Summerfield & R. Meddis), pp. 595–603. London, UK: Whurr Publishers.

Deng, L. & Geisler, C. D. 1987*a* A composite auditory model for processing speech sounds. *J. Acoust. Soc. Am.* **82**, 2001–2012. (doi:10.1121/1.395644)

Deng, L. & Geisler, C. D. 1987*b* Responses of auditory-nerve fibers to nasal consonant–vowel syllables. *J. Acoust. Soc. Am.* **82**, 1977–1988. (doi:10.1121/1.395642)

Diehl, R. L. 2008 Acoustic and auditory phonetics: the adaptive design of speech sound systems. *Phil. Trans. R. Soc. B* **363**, 965–978. (doi:10.1098/rstb.2007.2153)

Doupe, A. J. & Konishi, M. 1991 Song-selective auditory circuits in the vocal control system of the zebra finch. *Proc. Natl Acad. Sci. USA* **88**, 11 339–11 343. (doi:10.1073/pnas.88.24.11339)

Eggermont, J. J. 1977 Compound action potential tuning curves in normal and pathological human ears. *J. Acoust. Soc. Am.* **62**, 1247–1251. (doi:10.1121/1.381639)

Eggermont, J. J. 1993 Wiener and Volterra analyses applied to the auditory system. *Hear. Res.* **66**, 177–201. (doi:10.1016/0378-5955(93)90139-R)

Eggermont, J. J. 1995 Representation of a voice onset time continuum in primary auditory cortex of the cat. *J. Acoust. Soc. Am.* **98**, 911–920. (doi:10.1121/1.413517)

Eggermont, J. J. 2001 Between sound and perception: reviewing the search for a neural code. *Hear. Res.* **157**, 1–42. (doi:10.1016/S0378-5955(01)00259-3)

Ehret, G. 1997 The auditory midbrain, a "shunting yard" of acoustical information processing. In *The central auditory system* (eds G. Ehret & R. Romand), pp. 259–316. New York, NY: Oxford University Press.

Ehrlich, D., Casseday, J. H. & Covey, E. 1997 Neural tuning to sound duration in the inferior colliculus of the big brown bat, *Eptesicus fuscus*. *J. Neurophysiol.* **77**, 2360–2372.

Evans, E. F. 1992 Comparisons of physiological and behavioural properties: auditory frequency selectivity. In *Auditory physiology and perception* (eds N. P. Cooper, Y. Cazals & K. Horner), pp. 159–162. Oxford, UK: Pergamon.

Fant, G. 1970 *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton.

Fay, R. R. 1988 *Hearing in vertebrates: a psychophysics databook*. Winnetka, IL: Hill-Fay Associates.

Fritz, J., Shamma, S., Elhilali, M. & Klein, D. 2003 Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* **6**, 1216–1223. (doi:10.1038/nn1141)

Fuzessery, Z. M. & Hall, J. C. 1999 Sound duration selectivity in the pallid bat inferior colliculus. *Hear. Res.* **137**, 137–154. (doi:10.1016/S0378-5955(99)00133-1)

Gaese, B. H. & Ostwald, J. 2001 Anesthesia changes frequency tuning of neurons in the rat primary auditory cortex. *J. Neurophysiol.* **86**, 1062–1066.

Geisler, C. D. 1989 The responses of models of "high-spontaneous" auditory-nerve fibers in a damaged cochlea to speech syllables in noise. *J. Acoust. Soc. Am.* **86**, 2192–2205. (doi:10.1121/1.398480)

Geisler, C. D. & Gamble, T. 1989 Responses of "high-spontaneous" auditory-nerve fibers to consonant–vowel syllables in noise. *J. Acoust. Soc. Am.* **85**, 1639–1652. (doi:10.1121/1.397952)

Goldberg, J. M. & Brown, P. B. 1969 Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J. Neurophysiol.* **32**, 613–636.

Grace, J. A., Amin, N., Singh, N. C. & Theunissen, F. E. 2003 Selectivity for conspecific song in the zebra finch auditory forebrain. *J. Neurophysiol.* **89**, 472–487. (doi:10.1152/jn.00088.2002)

Green, D. M. & Swets, J. A. 1966 *Signal detection theory and psychophysics*. New York, NY: Wiley.

Greenwood, D. D. 1961 Critical bandwidth and the frequency coordinates of the basilar membrane. *J. Acoust. Soc. Am.* **33**, 1344–1356. (doi:10.1121/1.1908437)

Greenwood, D. D. 1990 A cochlear frequency-position function for several species—29 years later. *J. Acoust. Soc. Am.* **87**, 2592–2605. (doi:10.1121/1.399052)

Harrison, R. V., Aran, J. M. & Erre, J.-P. 1981 AP tuning curves from normal and pathological human and guinea pig cochleas. *J. Acoust. Soc. Am.* **69**, 1374–1385. (doi:10.1121/1.385819)

Hienz, R. D., Aleszczyk, C. M. & May, B. J. 1996 Vowel discrimination in cats: thresholds for the detection of second formant changes in the vowel /ɛ/. *J. Acoust. Soc. Am.* **100**, 1052–1058. (doi:10.1121/1.416291)

Holmes, S. D., Sumner, C. J., O'Mard, L. P. & Meddis, R. 2004 The temporal representation of speech in a nonlinear model of the guinea pig cochlea. *J. Acoust. Soc. Am.* **116**, 3534–3545. (doi:10.1121/1.1815111)

Irvine, D. R. F. 1992 Physiology of the auditory brainstem. In *The mammalian auditory pathway: neurophysiology* (eds A. N. Popper & R. R. Fay), pp. 153–231. New York, NY: Springer.

Javel, E. 1981 Suppression of auditory nerve responses I: temporal analysis, intensity effects and suppression contours. *J. Acoust. Soc. Am.* **69**, 1735–1745. (doi:10.1121/1.385953)

Javel, E., Geisler, C. D. & Ravindran, A. 1978 Two-tone suppression in auditory nerve of the cat: rate-intensity and temporal analyses. *J. Acoust. Soc. Am.* **63**, 1093–1104. (doi:10.1121/1.381817)

Johnson, D. H. 1980*a* Applicability of white-noise nonlinear system analysis to the peripheral auditory system. *J. Acoust. Soc. Am.* **68**, 876–884. (doi:10.1121/1.384826)

Johnson, D. H. 1980*b* The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J. Acoust. Soc. Am.* **68**, 1115–1122. (doi:10.1121/1.384982)

Johnson, D. H., Gruner, C. M., Baggerly, K. & Seshagiri, C. 2001 Information–theoretic analysis of the neural code. *J. Comput. Neurosci.* **10**, 47–69. (doi:10.1023/A:1008968010214)

Joris, P. X., Schreiner, C. E. & Rees, A. 2004 Neural processing of amplitude-modulated sounds. *Physiol. Rev.* **84**, 541–577. (doi:10.1152/physrev.00029.2003)

Kaas, J. H., Hackett, T. A. & Tramo, M. J. 1999 Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.* **9**, 164–170. (doi:10.1016/S0959-4388(99)80022-1)

Kandel, E. R., Schwartz, J. H. & Jessell, T. M. 2000 *Principles of neural science*. New York, NY: McGraw-Hill.

Keilson, S. E., Richards, V. M., Wyman, B. T. & Young, E. D. 1997 The representation of concurrent vowels in the cat anesthetized ventral cochlear nucleus: evidence for a periodicity-tagged spectral representation. *J. Acoust. Soc. Am.* **102**, 1056–1071. (doi:10.1121/1.419859)

Kiang, N. Y. S., Watanabe, T., Thomas, E. C. & Clark, L. F. 1965 *Discharge patterns of single fibers in the cat's auditory nerve*. Cambridge, MA: MIT Press.

Kiefte, M. & Kluender, K. R. 2001 Synthetic speech stimuli spectrally normalized for nonhuman cochlear dimensions. *Acoust. Res. Lett. Online* **3**, 41–46. (doi:10.1121/1.1445202)

Kim, D. O., Siegel, J. H. & Molnar, C. E. 1979 Cochlear nonlinear phenomena in two-tone responses. *Scand. Audiol.* **9**(Suppl.), 63–81.

Klein, D. J., Depireux, D. A., Simon, J. Z. & Shamma, S. A. 2000 Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J. Comput. Neurosci.* **9**, 85–111. (doi:10.1023/A:1008990412183)

Kuwada, S., Batra, R. & Stanford, T. R. 1989 Monaural and binaural response properties of neurons in the inferior colliculus of the rabbit: effects of sodium pentobarbital. *J. Neurophysiol.* **61**, 269–282.

Lai, Y. C., Winslow, R. L. & Sachs, M. B. 1994 The functional role of excitatory and inhibitory interactions in chopper cells of the anteroventral cochlear nucleus. *Neural Comput.* **6**, 1127–1140.

Lewis, E. R. & Henry, K. R. 1994 Dynamic changes in tuning in the gerbil cochlea. *Hear. Res.* **79**, 183–189. (doi:10.1016/0378-5955(94)90139-2)

Liberman, M. C. 1978 Auditory-nerve response from cats raised in a low-noise chamber. *J. Acoust. Soc. Am.* **63**, 442–455. (doi:10.1121/1.381736)

Liberman, M. C. 1980 Morphological differences among radial afferent fibers in the cat cochlea: an electron-microscopic study of serial sections. *Hear. Res.* **3**, 45–63. (doi:10.1016/0378-5955(80)90007-6)

Liberman, M. C. & Beil, D. G. 1979 Hair cell condition and auditory nerve response in normal and noise-damaged cochleas. *Acta Otolaryngol.* **88**, 161–176.

Liberman, M. C. & Dodds, L. W. 1984 Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves. *Hear. Res.* **16**, 55–74. (doi:10.1016/0378-5955(84)90025-X)

Liu, C. & Kewley-Port, D. 2004 Formant discrimination in noise for isolated vowels. *J. Acoust. Soc. Am.* **116**, 3119–3129. (doi:10.1121/1.1802671)

Machens, C. K., Wehr, M. S. & Zador, A. M. 2004 Linearity of cortical receptive fields measured with natural sounds. *J. Neurosci.* **24**, 1089–1100. (doi:10.1523/JNEUROSCI.4445-03.2004)

Margoliash, D. 1986 Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J. Neurosci.* **6**, 1643–1661.

May, B. J., Huang, A., Le Prell, G. & Hienz, R. D. 1996 Vowel formant frequency discrimination in cats: comparison of auditory nerve representations and psychophysical thresholds. *Aud. Neurosci.* **3**, 135–162.

May, B. J., LePrell, G. S., Hienz, R. D. & Sachs, M. B. 1997 Speech representation in the auditory nerve and ventral cochlear nucleus. In *Acoustical signal processing in the central auditory system* (ed. J. Syka), pp. 413–429. New York, NY: Plenum Press.

May, B. J., LePrell, G. S. & Sachs, M. B. 1998 Vowel representations in the ventral cochlear nucleus of the cat: effects of level, background noise, and behavioral state. *J. Neurophysiol.* **79**, 1755–1767.

Metherate, R., Kaur, S., Kawai, H., Lazar, R., Liang, K. & Rose, H. J. 2005 Spectral integration in auditory cortex: mechanisms and modulation. *Hear. Res.* **206**, 146–158. (doi:10.1016/j.heares.2005.01.014)

Miller, M. I. & Sachs, M. B. 1983 Representation of stop consonants in the discharge patterns of auditory-nerve fibers. *J. Acoust. Soc. Am.* **74**, 502–517. (doi:10.1121/1.389816)

Miller, R. L., Schilling, J. R., Franck, K. R. & Young, E. D. 1997 Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers. *J. Acoust. Soc. Am.* **101**, 3602–3616. (doi:10.1121/1.418321)

Miller, R. L., Calhoun, B. M. & Young, E. D. 1999*a* Discriminability of vowel representations in cat auditory-nerve fibers after acoustic trauma. *J. Acoust. Soc. Am.* **105**, 311–325. (doi:10.1121/1.424552)

Miller, R. L., Calhoun, B. M. & Young, E. D. 1999*b* Contrast enhancement improves the representation of /ɛ/-like vowels in the hearing-impaired auditory nerve. *J. Acoust. Soc. Am.* **106**, 2693–2708. (doi:10.1121/1.428135)

Mittmann, D. H. & Wenstrup, J. J. 1995 Combination-sensitive neurons in the inferior colliculus. *Hear. Res.* **90**, 185–191. (doi:10.1016/0378-5955(95)00164-X)

Møller, A. R. 1977 Frequency selectivity of single auditory-nerve fibers in response to broadband noise stimuli. *J. Acoust. Soc. Am.* **62**, 135–142. (doi:10.1121/1.381495)

Moore, B. C. J. 1995 *Perceptual consequences of cochlear damage*. Oxford, UK: Oxford University Press.

Moore, B. C. J. 2003 *An introduction to the psychology of hearing*. Amsterdam, The Netherlands: Elsevier.

Moore, B. C. J. 2008 Basic auditory processes involved in the analysis of speech sounds. *Phil. Trans. R. Soc. B* **363**, 947–963. (doi:10.1098/rstb.2007.2152)

Nelken, I. 2004 Processing of complex stimuli and natural scenes in the auditory cortex. *Curr. Opin. Neurobiol.* **14**, 474–480. (doi:10.1016/j.conb.2004.06.005)

Nelken, I., Fishbach, A., Las, L., Ulanovsky, N. & Farkas, D. 2003 Primary auditory cortex of cats: feature detection or something else? *Biol. Cybern.* **89**, 397–406. (doi:10.1007/s00422-003-0445-3)

O'Neill, W. E., Frisina, R. D. & Gooler, D. M. 1989 Functional organization of mustached bat inferior colliculus: I. Representation of FM frequency bands important for target ranging revealed by 14C-2-deoxyglucose autoradiography and single unit mapping. *J. Comp. Neurol.* **284**, 60–84. (doi:10.1002/cne.902840106)

Oliver, D. L. & Huerta, M. F. 1992 Inferior and superior colliculi. In *The mammalian auditory pathway: neuroanatomy* (eds D. B. Webster, A. N. Popper & R. R. Fay), pp. 168–221. New York, NY: Springer.

Oxenham, A. J. & Shera, C. A. 2003 Estimates of human cochlear tuning at low levels using forward and simultaneous masking. *J. Assoc. Res. Otolaryngol.* **4**, 541–554. (doi:10.1007/s10162-002-3058-y)

Palmer, A. R. 1990 The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns

of guinea pig cochlear-nerve fibers. *J. Acoust. Soc. Am.* **88**, 1412–1426. (doi:10.1121/1.400329)

Palmer, A. R. & Moorjani, P. A. 1993 Responses to speech signals in the normal and pathological peripheral auditory system. *Prog. Brain Res.* **97**, 107–115.

Palmer, A. R. & Russell, I. J. 1986 Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hear. Res.* **24**, 1–15. (doi:10.1016/0378-5955(86)90002-X)

Palmer, A. R., Winter, I. M. & Darwin, C. J. 1986 The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primary-like cochlear nucleus neurons. *J. Acoust. Soc. Am.* **79**, 100–113. (doi:10.1121/1.393633)

Pfeiffer, R. R. & Kim, D. O. 1975 Cochlear nerve fiber responses: distribution along the cochlear partition. *J. Acoust. Soc. Am.* **58**, 867–869. (doi:10.1121/1.380735)

Pickles, J. O. 1979 Psychophysical frequency resolution in the cat as determined by simultaneous masking and its relation to auditory-nerve resolution. *J. Acoust. Soc. Am.* **66**, 1725–1732. (doi:10.1121/1.383645)

Pickles, J. O. 1988 *An introduction to the physiology of hearing*. San Diego, CA: Academic Press.

Pinheiro, A. D., Wu, M. & Jen, P. H. 1991 Encoding repetition rate and duration in the inferior colliculus of the big brown bat, *Eptesicus fuscus*. *J. Comp. Physiol. A* **169**, 69–85. (doi:10.1007/BF00198174)

Ramachandran, R., Davis, K. A. & May, B. J. 1999 Single-unit responses in the inferior colliculus of decerebrate cats I. Classification based on frequency response maps. *J. Neurophysiol.* **82**, 152–163.

Reale, R. A. & Geisler, C. D. 1980 Auditory-nerve fiber encoding of two-tone approximations to steady-state vowels. *J. Acoust. Soc. Am.* **67**, 891–902. (doi:10.1121/1.383969)

Recio, A. & Rhode, W. S. 2000 Representation of vowel stimuli in the ventral cochlear nucleus of the chinchilla. *Hear. Res.* **146**, 167–184. (doi:10.1016/S0378-5955(00)00111-8)

Recio, A., Rhode, W. S., Kiefte, M. & Kluender, K. R. 2002 Responses to cochlear normalized speech stimuli in the auditory nerve of cat. *J. Acoust. Soc. Am.* **111**, 2213–2218. (doi:10.1121/1.1468878)

Recio-Spinoso, A., Temchin, A. N., van Dijk, P., Fan, Y. H. & Ruggero, M. A. 2005 Wiener–kernel analysis of responses to noise of chinchilla auditory-nerve fibers. *J. Neurophysiol.* **93**, 3615–3634. (doi:10.1152/jn.00882.2004)

Rhode, W. S. & Greenberg, S. 1992 Physiology of the cochlear nucleus. In *The mammalian auditory pathway: neurophysiology* (eds A. N. Popper & R. R. Fay), pp. 94–152. Berlin, Germany: Springer.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. & Bialek, W. 1997 *Spikes, exploring the neural code*. Cambridge, MA: MIT Press.

Robles, L. & Ruggero, M. A. 2001 Mechanics of the mammalian cochlea. *Physiol. Rev.* **81**, 1305–1352.

Romand, R. & Avan, P. 1997 Anatomical and functional aspects of the cochlear nucleus. In *The central auditory system* (eds G. Ehret & R. Romand), pp. 97–191. New York, NY: Oxford University Press.

Rose, J. E., Brugge, J. F., Anderson, D. J. & Hind, J. E. 1967 Phase-locked response to low frequency tones in single auditory nerve fibers of the squirrel monkey. *J. Neurophysiol.* **30**, 769–793.

Rosen, S. 1992 Temporal information in speech: acoustic, auditory and linguistic aspects. *Phil. Trans. R. Soc. B* **336**, 367–373. (doi:10.1098/rstb.1992.0070)

Rouiller, E. M. 1997 Functional organization of the auditory pathways. In *The central auditory system* (eds G. Ehret & R. Romand), pp. 3–96. New York, NY: Oxford University Press.

Ruggero, M. A. & Temchin, A. N. 2005 Unexceptional sharpness of frequency tuning in the human cochlea. *Proc. Natl Acad. Sci. USA* **102**, 18 614–18 619. (doi:10.1073/pnas.0509323102)

Ruggero, M. A., Rich, N. C., Recio, A. & Narayan, S. S. 1997 Basilar-membrane responses to tones at the base of the chinchilla cochlea. *J. Acoust. Soc. Am.* **101**, 2151–2163. (doi:10.1121/1.418265)

Sachs, M. B. 1969 Stimulus–response relation for auditory–noise fibers: two-tone stimuli. *J. Acoust. Soc. Am.* **45**, 1025–1036. (doi:10.1121/1.1911493)

Sachs, M. B. & Abbas, P. J. 1974 Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. *J. Acoust. Soc. Am.* **56**, 1835–1847. (doi:10.1121/1.1903521)

Sachs, M. B. & Young, E. D. 1979 Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *J. Acoust. Soc. Am.* **66**, 470–479. (doi:10.1121/1.383098)

Sachs, M. B., Voigt, H. F. & Young, E. D. 1983 Auditory nerve representation of vowels in background noise. *J. Neurophysiol.* **50**, 27–45.

Salvi, R., Perry, J., Hamernik, R. P. & Henderson, D. 1982 Relationships between cochlear pathologies and auditory nerve and behavioral responses following acoustic trauma. In *New perspectives on noise-induced hearing loss* (eds R. P. Hamernik, D. Henderson & R. Salvi), pp. 165–188. New York, NY: Raven.

Schilling, J. R., Miller, R. L., Sachs, M. B. & Young, E. D. 1998 Frequency shaped amplification changes the neural representation of speech with noise-induced hearing loss. *Hear. Res.* **117**, 57–70. (doi:10.1016/S0378-5955(98)00003-3)

Schmiedt, R. A., Zwislocki, J. J. & Hamernik, R. P. 1980 Effects of hair cell lesions on responses of cochlear nerve fibers. I. Lesions, tuning curves, two-tone inhibition, and responses to trapezoidal-wave patterns. *J. Neurophysiol.* **43**, 1367–1389.

Schreiner, C. E. 1998 Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. *Audiol. Neuro-Otol.* **3**, 104–122. (doi:10.1159/000013785)

Schreiner, C. E., Read, H. L. & Sutter, M. L. 2000 Modular organization of frequency integration in primary auditory cortex. *Annu. Rev. Neurosci.* **23**, 501–529. (doi:10.1146/annurev.neuro.23.1.501)

Sen, K., Theunissen, F. E. & Doupe, A. J. 2001 Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.* **86**, 1445–1458.

Shamma, S. A. 1985 Speech processing in the auditory system. II: lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J. Acoust. Soc. Am.* **78**, 1622–1632. (doi:10.1121/1.392800)

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. & Ekelid, M. 1995 Speech recognition with primarily temporal cues. *Science* **270**, 303–304. (doi:10.1126/science.270.5234.303)

Shepherd, G. M. 2004 *The synaptic organization of the brain*. Oxford, UK: Oxford University Press.

Shera, C. A., Guinan, J. J. & Oxenham, A. J. 2002 Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc. Natl Acad. Sci. USA* **99**, 3318–3323. (doi:10.1073/pnas.032675099)

Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P. & Ruggero, M. A. 2005 Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering. *J. Acoust. Soc. Am.* **118**, 2434–2443. (doi:10.1121/1.2005867)

Silkes, S. M. & Geisler, C. D. 1991 Responses of "lower-spontaneous-rate" auditory-nerve fibers to speech syllables presented in noise. I: general characteristics. *J. Acoust. Soc. Am.* **90**, 3122–3139. (doi:10.1121/1.401421)

Sinex, D. G. 1993 Auditory nerve fiber representation of cues to voicing in syllable-final stop consonants. *J. Acoust. Soc. Am.* **94**, 1351–1362. (doi:10.1121/1.408163)

Sinex, D. G. & Chen, G.-D. 2000 Neural responses to the onset of voicing are unrelated to other measures of temporal resolution. *J. Acoust. Soc. Am.* **107**, 486–495. (doi:10.1121/1.428316)

Sinex, D. G. & Geisler, C. D. 1983 Responses of auditory-nerve fibers to consonant–vowel syllables. *J. Acoust. Soc. Am.* **73**, 602–615. (doi:10.1121/1.389007)

Sinex, D. G. & Geisler, C. D. 1984 Comparison of the responses of auditory nerve fibers to consonant–vowel syllables with predictions from linear models. *J. Acoust. Soc. Am.* **76**, 116–121. (doi:10.1121/1.391106)

Sinex, D. G. & McDonald, L. P. 1988 Average discharge rate representation of voice onset time in the chinchilla auditory nerve. *J. Acoust. Soc. Am.* **83**, 1817–1827. (doi:10.1121/1.396516)

Sinex, D. G. & Narayan, S. S. 1994 Auditory-nerve fiber representation of temporal cues to voicing in word-medial stop consonants. *J. Acoust. Soc. Am.* **95**, 897–903. (doi:10.1121/1.408400)

Sinex, D. G., McDonald, L. P. & Mott, J. B. 1991 Neural correlates of nonmonotonic temporal acuity for voice onset time. *J. Acoust. Soc. Am.* **90**, 2441–2449. (doi:10.1121/1.402048)

Smith, R. L. 1977 Short-term adaptation in single auditory nerve fibers: some poststimulatory effects. *J. Neurophysiol.* **40**, 1098–1112.

Smith, R. L. 1979 Adaptation, saturation and physiological masking in single auditory-nerve fibers. *J. Acoust. Soc. Am.* **65**, 166–178. (doi:10.1121/1.382260)

Smits, R., ten Bosch, L. & Collier, R. 1996 Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment. *J. Acoust. Soc. Am.* **100**, 3852–3864. (doi:10.1121/1.417241)

Steinschneider, M., Schroeder, C. E., Arezzo, J. C. & Vaughan Jr, H. G. 1995 Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: temporal response patterns. *Brain Lang.* **48**, 326–340. (doi:10.1006/brln.1995.1015)

Steinschneider, M., Fishman, Y. I. & Arezzo, J. C. 2003 Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *J. Acoust. Soc. Am.* **114**, 307–321. (doi:10.1121/1.1582449)

Stevens, K. N. & Blumstein, S. E. 1978 Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am.* **64**, 1358–1368. (doi:10.1121/1.382102)

Stevens, K. N. & House, A. S. 1961 An acoustical theory of vowel production. *J. Speech Hear. Res.* **4**, 303–320.

Theunissen, F. E., David, S. V., Singh, N. C., Hsu, A., Vinje, W. E. & Gallant, J. L. 2001 Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network Comput. Neural Syst.* **12**, 289–316. (doi:10.1088/0954-898X/12/3/304)

Ulanovsky, N., Las, L. & Nelken, I. 2003 Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* **6**, 391–398. (doi:10.1038/nn1032)

Van Tasell, D. J., Soli, S. D., Kirby, V. M. & Widin, G. P. 1987 Speech waveform envelope cues for consonant recognition. *J. Acoust. Soc. Am.* **82**, 1152–1161. (doi:10.1121/1.395251)

Versnel, H. & Shamma, S. A. 1998 Spectral-ripple representation of steady-state vowels in primary auditory cortex. *J. Acoust. Soc. Am.* **103**, 2502–2514. (doi:10.1121/1.422771)

Wang, X. 2000 On cortical coding of vocal communication sounds in primates. *Proc. Natl Acad. Sci. USA* **97**, 11 843–11 849. (doi:10.1073/pnas.97.22.11843)

Wang, X. & Kadia, S. C. 2001 Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J. Neurophysiol.* **86**, 2616–2620.

Wang, X., Merzenich, M. M., Beitel, R. & Schreiner, C. E. 1995 Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J. Neurophysiol.* **74**, 2685–2706.

Winer, J. A. & Schreiner, C. E. 2005 *The inferior colliculus*. New York, NY: Springer.

Winter, I. M. & Palmer, A. R. 1990 Temporal responses of primarylike anteroventral cochlear nucleus units to the steady-state vowel /i/. *J. Acoust. Soc. Am.* **88**, 1437–1441. (doi:10.1121/1.399720)

Winter, I. M., Robertson, D. & Yates, G. K. 1990 Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibres. *Hear. Res.* **45**, 191–202. (doi:10.1016/0378-5955(90)90120-E)

Wong, S. W. & Schreiner, C. E. 2003 Representation of CV-sounds in cat primary auditory cortex: intensity dependence. *Speech Commun.* **41**, 93–106. (doi:10.1016/S0167-6393(02)00096-1)

Wong, J. C., Miller, R. L., Calhoun, B. M., Sachs, M. B. & Young, E. D. 1998 Effects of high sound levels on responses to the vowel /ɛ/ in cat auditory nerve. *Hear. Res.* **123**, 61–77. (doi:10.1016/S0378-5955(98)00098-7)

Yates, G. K., Winter, I. M. & Robertson, D. 1990 Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. *Hear. Res.* **45**, 203–220. (doi:10.1016/0378-5955(90)90121-5)

Yeshurun, Y., Wollberg, Z. & Dyn, N. 1989 Prediction of linear and non-linear responses of MGB neurons by system identification methods. *Bull. Math. Biol.* **51**, 337–346.

Yin, T. C. T. 2002 Neural mechanisms of encoding binaural localization cues in the auditory brainstem. In *Integrative functions in the mammalian auditory pathway* (eds D. Oertel, A. N. Popper & R. R. Fay), pp. 99–159. New York, NY: Springer.

Yin, T. C. T. & Chan, J. C. K. 1990 Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.* **64**, 465–488.

Young, E. D. & Oertel, D. 2003 The cochlear nucleus. In *Synaptic organization of the brain* (ed G. M. Shepherd), pp. 125–163. New York, NY: Oxford University Press.

Young, E. D. & Sachs, M. B. 1979 Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J. Acoust. Soc. Am.* **66**, 1381–1403. (doi:10.1121/1.383532)

Young, E. D. & Sachs, M. B. 1981 Processing of speech in the peripheral auditory system. In *The cognitive representation of speech* (eds T. Myers, J. Laver & J. Anderson), pp. 75–92. Amsterdam, The Netherlands: North-Holland.

Zurita, P., Villa, A. E., de Ribaupierre, Y., de Ribaupierre, F. & Rouiller, E. M. 1994 Changes of single unit activity in the cat's auditory thalamus and cortex associated to different anesthetic conditions. *Neurosci. Res.* **19**, 303–316. (doi:10.1016/0168-0102(94)90043-4)